Vol. 67, No. 5 September 1960

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

CONTENTS

ARTICLES

Discrimination and Learning without Awareness: A Methodological Survey and Evaluation.....CHARLES W. ERIKSEN 279

Some Implications of a Social Learning Theory for the Prediction of Goal Directed Behavior from Testing Procedures
JULIAN B. ROTTER 301

Learning Theory and the Acquisition of Values..WINFRED F. HILL 317

Word Values, Word Frequency, and Visual Duration Thresholds
RONALD C. JOHNSON, CALVIN W. THOMSON, AND GERALD FRINCKE 332

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Editorial Consultants

JOE K. ADAMS
FLOYD H. ALLPORT
JOHN W. ATKINSON
FRANK A. BEACH
JEROME S. BRUNER
ARTHUR S. COUCH
PAUL FITTS
FREDERICK FRICK
EUGENE H. GALANTER
WENDELL R. GARNER
JAMES GIBSON
RICHARD HERRNSTEIN
DAVIS HOWES
J. MCV. HUNT
WILLIAM LAMBERT
THOMAS LANDAUER
GARDNER LINDZEY
ELEANOR E. MACCOBY
RICHARD MANN
PAUL MEEHL
DONALD R. MEYER
BENNET MURDOCK
CHARLES E. OSGOOD
RICHARD K. OVERTON
J. TREVOR PEIRCE
FRANK RESTLE
BURTON ROSNER
HAROLD SCHLOSBERG
K. U. SMITH
KENNETH SPENCE
JOHN P. SPIEGEL
ELIOT STELLAR
S. S. STEVENS
FREDERICK STROTBECK
HENRY TAJFEL
WARREN TORGERSON
BENTON J. UNDERWOOD
WILLIAM S. VERPLANCK
JOHN WHITING

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals. However, reports of original research may be included in a major theoretical article which attempts to integrate several related original studies.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted.

All manuscripts must be submitted in duplicate.

Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript, in case there is loss in the mail. Manuscripts should be addressed to the Editor, Richard L. Solomon, Psychology Department, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania.

ARTHUR C. HOFFMAN
Managing Editor

HELEN CRE
Promotion Manager

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1333 Sixteenth St. N. W., Washington 6, D. C.

$8.00 volume $1.50 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

VOL. 67, NO. 5 SEPTEMBER 1960

THE PSYCHOLOGICAL REVIEW

DISCRIMINATION AND LEARNING WITHOUT AWARENESS:

A METHODOLOGICAL SURVEY AND EVALUATION [1]

CHARLES W. ERIKSEN

University of Illinois

The importance of a concept of the unconscious in psychology does not need to be documented. Most, if not all, current theories of personality employ the concept in their treatment of such diverse problems as defense mechanisms, motivation, psychopathology, and psychotherapy. Despite the importance of this concept little or no progress has been made in refining and objectifying it. The current treatment of the unconscious is little if any more advanced than the loose animistic concept devised by Freud. This looseness gives rise to excessive surplus meaning and many powers (occasionally bordering on the mystical) are glibly and uncritically attributed to this pervasive process.

While there has been a resurgence of experimental interest in unconscious phenomena, this research for the most part tends to be primarily demonstrational in nature rather than parametric. The complexity of the experimental operations affords numerous opportunities for artifacts to occur and the results are generally overinterpreted. In this latter connection little or no attempts are made to relate the phenomena to the general body of experimentally derived psychological knowledge.

This paper presents a representative survey of experimental approaches to unconscious phenomena with the purpose of examining methodological and definitional problems that are encountered. An attempt is made to separate different phenomena that have been lumped together under the general concept of the unconscious and to treat these phenomena in terms of current laboratory derived concepts.

PROBLEMS OF DEFINITIONS

The nonscientific nature of current conceptions of the unconscious is attested by the vagueness of definitions of what constitutes conscious and unconscious behavior. Not only are the definitions vague and operationally inadequate but there are several different terms that are used, sometimes to designate the same phenomena, other times in a partially nonoverlapping way. Such terms as "consciousness," "awareness," "insight," and "attention" have all been used to describe the same phenomena. I do not intend to trace the differences or the commonalities among these different terms and their application to specific problems but

[1] This paper was facilitated by Research Grant M-1206, National Institute of Mental Health, Public Health Service.

merely to indicate the nature of the confusion that exists. In the remainder of the paper I shall use the terms awareness and unawareness to designate what is generally subsumed under the psychoanalytic terms, conscious and unconscious. I prefer the term awareness since it seems to reflect more closely the psychological quality of the phenomena whereas consciousness seems more appropriate for the designation of the state of physiological alertness of an organism.

The most frequent definition of awareness and unawareness is in terms of verbal report. Awareness is equated with the ability to verbalize and unawareness with the inability or lack of verbalization. Freud (1925) in his extensive writings can be found on occasion to have defined awareness and unawareness in this manner. Such a distinction has been popular with other writers on the subject with Dollard and Miller (1950) most explicit. They define unawareness as the unverbalized or the unlabeled and their distinction between aware and unaware processes follows from their conception of the role of verbalization in the individual's adjustment. Verbal labels are assumed to increase the precision of discrimination and to mediate the transfer of training through the cue producing characteristics of the labels.

The popularity of a definition of awareness in terms of verbal report is understandable. It creates the illusion of operational precision and helps to mask the fact that many of the more important characteristics of awareness are avoided by the definition. Such an equation might seem a simple solution to the definitional problem but it is to be noted that the operational meaning of such a definition in any given experimental situation hinges critically upon a number of considerations. Among these we need to consider the adequacy of the questioning of the subject (S), the motivation of the S to respond with the care and precision that is required, the care taken to assure that the S understands what is being asked him, consideration of the effects of the interrogation itself upon the delicate process of awareness, and most importantly an adequate schema for classifying the S's verbalizations along relevant dimensions. Assuming the last criterion can be met, a further requirement is some system of scaling or grading the S's verbalizations along these dimensions in terms of their accuracy or specificity. It is not sufficient, as is too often the case, for the experimenter (E) to make a general statement that interrogation of his Ss revealed that they were unaware. It is the E's decisions with respect to the above considerations that constitute the adequacy of his operational definition of awareness and any value the concept has depends upon the similarity of these operations from one experiment to another.

While attention to the above requirements constitutes a minimum standard for a definition of unawareness, it also is to be noted that there are some basic objections to the equation of awareness with a verbal report. There is nothing profound in noting that language and words are highly abstract symbols and bear little or no physical resemblance to the objects, events, and relationships that they denote. A definition of awareness in terms of verbalization places a heavy burden upon the adequacy of language to reflect the richness of perceptual experiences and images. The equation of awareness with verbal report leads to a similar kind of confounding of response and concept that has occurred with perception and the responses from which perception is inferred (Garner, Hake, & Eriksen, 1956).

There is also the question of how such a definition would handle voluntary responses other than verbal report. Thus in an experiment on complex perceptual discrimination an *S* may be at a loss to describe verbally the stimulus that has been presented to him but he can accurately choose the correct stimulus in a multiple-choice situation. Similarly, one's verbal description of the facial appearance of another individual may be wholly inadequate in enabling another to recognize this person among a crowd. However, if the person giving the description possessed a mediocre artistic talent, it would be possible for him to draw a sketch of the person described which would greatly increase the chances of the individual's being recognized. In these situations, if we adhered strictly to a definition of awareness in terms of verbalization, it would be necessary to say that the *S* was not aware of what the complex perceptual stimulus was like or of the described person's appearance. However, most of us will feel that this is failing to capture what we mean when we describe lack of awareness.

An equation of awareness with verbal report also would exclude from consideration a large body of experimental data that has supposedly dealt with the problem of awareness. As Adams (1957) has pointed out, one of the most readily demonstratable experimental arrangements supposedly showing unawareness is the following. In a psychophysical threshold determination if the *S* is instructed to give a confidence rating along with his judgments, it is frequently found that judgments which the *S* reports as pure guesses show above chance discrimination. If we equate awareness with verbal reportability, however, we would have to classify this as discrimination with awareness since the *S* is making a discriminated verbal report.

I have raised these objections and problems concerning the definition of awareness, not because I have a solution for them or a better definition to offer, but solely to insure that the issues raised in the remainder of this paper will be evaluated in terms of the limitations of our current definitions. I believe that in view of our present knowledge in this relatively unexplored research area, a definition of awareness in terms of a verbal report is a beginning point. But as was pointed out above any value such a definition has depends critically upon the operational specifications used in assessing the *S*s' verbalizations or lack thereof.

DISCRIMINATION WITHOUT AWARENESS

The question of discrimination without awareness is two separate questions with quite different implications. When we talk about discrimination without awareness we might mean: (*a*) is there a more sensitive (accurate) discrimination that occurs at levels not available to awareness; or (*b*) are we aware of all the discriminations that we make. The first question concerns the possibility that our nervous system differentially responds to stimuli or cues whose intensity is below that where a differentiated or discriminated verbal report can be obtained (subliminal discrimination). The second question concerns the possibility of differential responses to stimuli or cues above thereshold but which upon questioning the *S* is unable to state that he used these particular stimuli as cues for his responses (incidental discrimination). These two questions are seldom distinguished and it is not uncommon to find a psychologist arguing the affirmative answer to the first question

by citing experimental data that only demonstrate the latter.

When discrimination at levels below that available to awareness is explicitly considered, it is often assumed that the discrimination occurring at these levels is only of a rather gross type. Cason and Katcher (1933) give a very plausible neurophysiological speculation in terms of how such subliminal stimuli could operate. They state:

> Before a conscious activity comes into existence, the energy that is produced or liberated in the body must be a certain required strength. The nervous energy frequently leads to some bodily effect even though its strength is not sufficient to produce conscious activity. It seems reasonable to believe that energy of this subliminal strength can be connected with some response by the learning process and if this could be demonstrated the theoretical implications would be important (p. 831).

Lazarus (1956) in defending his interpretation of the subception effect has suggested that discrimination at levels below that available to awareness as measured by the GSR, is capable of making gross discriminations between threatening and nonthreatening stimuli. Osgood (1957) has made a similar suggestion for interpreting perceptual sensitization and defense data. He conceives that an overall affective evaluation occurs early in the perceptual process and

> that hints provided by the incomplete information in the stimulus display, by eliciting the dominant evaluative (emotional) components of mediating reaction which feed cues back to the perceptual system, may produce either facilitation of covaluent integrations (vigilance) or disorganization of integrative mechanisms (defense).

He further suggests that this process may be mediated by lower brain centers.

Conceiving of discrimination at levels below awareness as only a rather gross type of discrimination has a certain plausibility to it. It avoids the concept of the "unconscious" as a sort of supernatural entity with mysterious powers. On cursory examination it would even appear to have experimental support. Binder (1955) among others, has pointed out that one or more of the characteristics of a stimulus may be discriminated in the absence of sufficient information of the remaining characteristics to permit specific identification. Thus, at a brief exposure we may be able to discriminate the color of a stimulus without perceiving its texture or its shape, and, if we can discriminate the color of a stimulus without also perceiving the texture and shape necessary for identification, why couldn't we be expected to discriminate the affective value of a stimulus in the absence of a specific identification?

The difficulty with such a conclusion is this. Generally, the *meaning* of a stimulus is conveyed not by one or two physical characteristics but by the total aggregrate of characteristics. In the case of words, which are frequently used as stimuli in perceptual defense studies, the affective meaning is conveyed only by the perception of a complete word, not by isolated characteristics such as word length or that there is a tall letter at the beginning of the word. To discriminate the word shot from a vulgar four letter Anglo-Saxon word would require a very precise discrimination as to a subtle form difference in only one of the four letters. For discrimination of the affective meaning of a threatening stimulus to occur, some place in the nervous system the complete word would have to be identified. So, we find that at what first seems to be a plausible assumption of only very gross discrimination at levels below awareness, is in fact an assumption of very precise discrimination, provided it is to be

used to explain such phenomena as perceptual defense.

Experimental Studies of Subliminal Discrimination

What experimental evidence is there of a capacity for discrimination at levels below that available to verbal report? One can distinguish four basic methodological approaches to this problem. The first consists of attempting to condition some "nonvoluntary" response such as pupillary contractions or GSR to a stimulus presented with an intensity that has been previously determined to be subliminal. The second approach is to condition a GSR or similar nonvoluntary response to a stimulus at supraliminal levels and then use this response to assess the threshold for the stimulus. The third methodology, which is sometimes combined with the other two is to condition the GSR or some nonverbal response (or assume prior conditioning of this response) to a stimulus that is then presented around threshold values. The nonverbal response is then observed to determine whether it shows greater than chance relationship with the stimulus on those trials when the verbal judgment was incorrect. The fourth methodology is to present a stimulus, frequently a complex one, at a level assumed to be subliminal and then to look for the effect of the stimulation in subsequent behavior such as dreams, images, or free associations. I shall not attempt to summarize here all the different studies using these methodologies but instead select representative experiments which illustrate the general findings and the methodological problems encountered that limit the interpretation of the data.

Subliminal Conditioning

The evidence for subliminal conditioning is for the most part negative or at best controversial. One of the more intriguing studies is that of Baker (1938) who reported that the pupillary light reflex could be conditioned to a tone below the absolute threshold. Using the below threshold tone as a conditioned stimulus (CS) and change in illumination as the unconditioned stimulus (US), he was successful in establishing a conditioned response to the subthreshold tone within three trials. In addition, responses conditioned to the subthreshold sound generalized less than responses conditioned to suprathreshold values. They were also more stable and resistant to extinction.

In a repetition of the study, Metzner and Baker (1939) obtained essentially the same results, but other investigators have been unsuccessful in attempts to duplicate this experiment (Hilgard, Miller, & Ohlson, 1941; Wedell, Taylor, & Skolnick, 1940). Wedell, Taylor, and Skolnick found no evidence of conditioning in any of the *S*s they employed and a number of careful investigators have been unable to obtain pupillary conditioning even to suprathreshold stimulus values (Hilgard, Dutton, & Helmick, 1949; Steckle, 1936; Steckle & Renshaw, 1934; Stern, 1948).

In a recent experiment Taylor (1953) claims to have demonstrated subthreshold conditioning, but here again this interpretation of his results is open to serious question. He presented six different geometric figures tachistoscopically one at a time to determine a level of exposure where the *S*s were unable to identify at better than at chance level which specific geometric figure was being presented. He was successful in obtaining GSR conditioning at an exposure level where the *S*s were unable to correctly identify the different geometric figures.

While Taylor interprets his data in

terms of subthreshold discrimination, there is a more simple and plausible explanation of his findings. While his *S*s may have been unable to identify which geometric form was being exposed, this in itself is insufficient evidence for assuming that all six geometric forms at this level of exposure resulted in the same perceptual experience. It is more than likely that different perceptual experiences were associated with each of the six geometric forms at the level of exposure selected by Taylor. This is not to say that the information the *S* received was sufficient to specify what form was being shown, but merely that different experiences or perceptions were associated with the different forms. For example, the exposure of a square might have led to a description on the part of the *S* of a perception of a dark smudge that seemed to sort of expand in the field. A free report of his perception following exposure of the triangle might have led to some description such as "I had the impression of sort of a slanted line moving across the field." While these perceptions would not be sufficient to enable the *S* to label which figure was being presented, they nonetheless are distinguishable perceptions and therefore capable of differential conditioning or even differential verbal response provided proper feedback is given.

In Taylor's experiment a necessary control would have been to determine whether the *S*s reported the tachistoscopic exposure of a square and a triangle as identical perceptual experiences or as different. By asking the *S* to say "same" or "different" following successive exposures of geometric figures, one could have determined whether differential perceptual experiences existed at the verbally distinguishable level.

The possibility of subliminal conditioning of finger retraction was thoroughly and carefully explored by Newhall and Sears (1933). Using electric shock as the US and a light as the CS, these investigators were successful in obtaining conditioned finger retraction to light intensities at the absolute threshold (defined as the 50% point using the method of constant stimuli). Threshold determinations were also obtained throughout the experiment. Some of their *S*s were found to give conditioned finger retractions to light intensities below the absolute threshold (four of their seven *S*s gave a total of five conditioned finger retractions in the sigma intervals of $-.5$ to -1.0).

Newhall and Sears were properly cautious in interpreting their data in terms of subliminal conditioning. They pointed out that while some conditioned finger retractions were obtained at values below the absolute threshold, the threshold in their study was defined as the 50% point. Thus by definition of the threshold, the *S*s verbally would report on a number of occasions having seen the light at values that actually fall below the arbitrarily defined threshold.

Further confirmation that there was no super discrimination by unaware process in their experiment is found when the conditioned response is itself used to estimate the absolute threshold for light. By using the latter sessions in the experiment when conditioning had been well established in most *S*s, it was possible to estimate the thresholds from the frequency of conditioned finger retractions. It is interesting that the absolute threshold computed from the frequency of conditioned responses corresponds almost exactly to the value obtained by using the verbal judgments.

The most recent and perhaps the most thorough investigation of sub-

liminal conditioning appears to be that of Wilcott (1953). He attempted to condition GSR to subthreshold auditory intensities at several different frequency levels. His *S*s were given preliminary training on threshold determination and the stimulus intensity selected for subthreshold conditioning was the *highest intensity* that the *S failed* to report hearing during the threshold determination sessions. While he was successful in GSR conditioning to suprathreshold intensities, no evidence was found for conditioned GSRs in response to the subthreshold intensities at any frequency level. Interestingly enough, Wilcott's experiment appears to be the only one using a subliminal CS that was at a low enough intensity so as not to yield better than chance verbal reports.

From this brief survey of the literature we can see that the evidence for discrimination at levels below those available to awareness is not very impressive. In the cases where studies have obtained positive results, other investigators have been unable to duplicate the findings. In the cases where slight evidence for subthreshold conditioning has been obtained the results seem to be more readily accountable by other explanations.

There are several sources of artifact that need to be carefully considered before interpreting the data of an experiment in terms of conditioning or discrimination at levels below those available to awareness. Of major importance is the definition of a threshold. It must be borne in mind that our thresholds are arbitrary statistical definitions. In the typical psychophysical study an *S* is asked to report "yes" or "no" concerning the presence or absence of the stimulus. Commonly the 50% accuracy level has been accepted as a threshold. It is to be noted, however, that this procedure insures that the *S* is discriminating better than chance at a stimulus intensity below the arbitrary threshold designation. If a blank stimulus interval is included in the threshold determination, the *S*'s *false alarm rate* or his tendency to say "yes" in the absence of any stimulation can be observed. In a good *S* this false alarm rate may be 10% or less. Thus, any deviation from this 10% point in terms of positive answers or reports that he perceives the stimulus indicates he is now beginning to perceive at better than a chance level. At a stimulus intensity that the *S* reports perceiving 25% of the time, we are below his arbitrarily defined threshold value but at the same time we are 15% above his chance performance level. To demonstrate that a conditioned response can be obtained 15% of the time at this stimulus intensity level is hardly evidence for concluding discrimination below verbal awareness.

Further sources of error in the interpretation of subliminal conditioning studies revolve around the method used in obtaining thresholds and the instructions and training given the *S*s. We shall not go into these problems in detail here since the interested reader can consult the excellent discussion of these problems by Goldiamond (1958). However, we may note that the "threshold" obtained for an *S* depends upon a number of variables. Most important are: (*a*) the instructions and expectations of the *S*s; (*b*) the penalties for false alarms and for failure to detect a real stimulus; (*c*) the training given the *S*; (*d*) the use or lack of use of a ready signal; and (*e*) whether or not a forced-choice method is employed (Blackwell, 1953). Most typically experiments on subliminal stimulation have given only cursory attention to the operations by which they have concluded that a given stimulus was "subliminal." Few, if any, would

meet the exacting requirements of modern psychophysics.

Assessing Thresholds by Conditioned Responses

A second method for investigating the question of a more sensitive discrimination that occurs at levels below awareness is to condition some response other than verbal to a stimulus and then observe whether this response generalizes to values of the stimulus at threshold intensities. One could condition a GSR to a suprathreshold stimulus and then observe the frequency of conditioned responses as the stimulus was varied in intensity around the threshold level. The frequency of conditioned responses as a function of the intensity of the stimulus could be used to compute a threshold that could then be compared with the threshold obtained by verbal report. The experiment by Newhall and Sears described above provided an example of this method of threshold determination.

One of the more interesting experiments of this type is reported by Davis (1950). Recordings were made of the action potentials in several different muscle groups in response to a sound presented above and below threshold. Increases in muscle tension occurred to stimulus intensities that the *S*s did not report hearing. Unfortunately, these data can be taken as only suggestive since the method of threshold determination followed by Davis was not very rigorous. The *S*s were not given preliminary training nor were they alerted for the intervals in which the stimulus might appear. In addition the instructions given were not of the type that generally yield a most sensitive verbal threshold.

Even if replication should substantiate Davis' finding that muscle tension varies at levels of stimulation below those capable of differential verbal report, there is still the question of what meaning this has for behavior. For responses by some part of the body to external stimulation to be important psychologically it needs to be shown that these responses are capable of learning and serving as cues for behaviorally more important responses.

Recently Dulany and Eriksen (1959) reported an experiment in which the GSR threshold for brightness was systematically compared with verbal response thresholds determined by two methods—a modification of Blackwell's forced-choice as well as the more typical presence or absence report from the *S*. *S*s were trained for 10 days prior to a conditioning session. The GSR was then conditioned to a suprathreshold intensity of the test light with electroshock to the ankle as the US. Following conditioning, 3 more daily sessions were employed to establish the GSR threshold as well as a concurrent forced-choice discrimination threshold. To make the thresholds obtainable from the GSR and the verbal response directly comparable, the test patch could occur in either of two intervals designated by an auditory signal. *S*s were required to report verbally which interval the light occurred in and their GSRs were concurrently recorded. If the GSR to the first interval was greater than the GSR to the second, this was considered the GSR's choice on that particular trial.

While these investigators were successful in obtaining stable GSR thresholds in their *S*s, there was no indication that the GSR gave a lower threshold than did either of the two verbal report methods. In fact, the verbal reports in most *S*s were found to show a deviation from the chance base line at lower stimulus intensity than the GSR's. Also, it was noted that while

the verbal reports obtained an accuracy level of 100% at higher stimulus intensities, the GSR never achieved this high level of discrimination. This latter observation is important in that any attempt to claim greater sensitivity for the GSR would have to assume that somehow a response which is less sensitive or discriminating at above threshold intensities metamorphoses over the stimulus range to become more sensitive at low stimulus intensities.

It is to be noted that the Dulany and Eriksen experiment does not prove that the GSR is no more sensitive than the verbal report in obtaining a discrimination threshold. It is quite possible that the many sources of error that occur in GSR recording whether due to the equipment or to the nature of the response itself may prevent a fair comparison of the two response systems. Also, there is a possibility that *S*s have had greater practice in using the verbal responses as opposed to the GSR for such a discrimination. These, at the present time, appear to be unanswerable possibilities but it is to be noted that the care taken in training GSR and verbal response thresholds in this experiment was much more extensive than that employed in experiments that have used GSRs as an indicator of unconscious perception.

The Subception Effect

There is a third class of experiments that is erroneously but nonetheless often considered to show discrimination at levels below that available to awareness. These experiments make use of the same experimental arrangement as that employed in the previous approaches. A nonvoluntary nonverbal response such as a GSR is conditioned to some suprathreshold stimulus or stimuli and then the stimuli are presented at about threshold value. The *S*s are required to give a verbal response and the incidence of the concurrent conditioned response is noted. However, in these experiments the conditioned response is not used to compute a threshold. Instead the data are analyzed in terms of whether or not a conditioned response occurs more often than chance would indicate in the absence of a correct verbal response. Lazarus and McCleary (1951) in their subception experiment were among the first to use this type of analysis as evidence for discrimination without awareness.

They conditioned GSRs to 5 of 10 nonsense syllables and then exposed the nonsense syllables individually in a tachistoscope at a duration level where verbal recognition was about 50% correct. The average GSR to shock syllables was found to be greater than to nonshock syllables on those trials where the *S*s were unable to give the correct verbal identification of the syllables. They interpreted their data in terms of subception: "a process by which some kind of discrimination is made when the *S* is unable to make a correct conscious discrimination."

While nearly all of the discrimination without awareness experiments that we have analyzed so far would have lent their data to this type of comparison, the novelty of the Lazarus and McCleary approach was the introduction of this method of analysis. It is not the same as determining a differential threshold for the GSR and the verbal response. Lazarus and McCleary did not compute separate thresholds. While they report that the tachistoscopic level was set so as to obtain approximately 50% correct verbal recongnitions, they do not report the differential accuracy of the GSR on classifying shock and nonshock nonsense syllables. Such information is important to our present discussion and

can be estimated from an unpublished replication of their study conducted in our laboratory. This replication was successful in duplicating the subception effect. But a comparison of the relative accuracy of the GSR and the verbal report shows that if an observer wished to predict whether a shock or a nonshock nonsense syllable had been presented to the *S* and the observer could use only one of the *S*'s responses, the verbal report or the GSR, he would have been appreciably more accurate using the *S*'s verbal report.

At the exposure level used in the replication approximately 45% of the verbal reports correctly identified the specific nonsense syllable. The most optimum weighting of GSR values gave approximately 60% correct classifications into shock or nonshock categories, but the chance expectancy for correct nonsense syllable identification for the verbal response is 10%, whereas for the dichotomous classification of shock vs. nonshock by the GSR, chance level is 50%. An observer attempting to determine whether the *S* had seen a shock or a nonshock syllable would have been correct in his predictions 60% of the time using the GSR but using the verbal response he would have been correct 45% of the time plus one-half of 55% for a total of 72.5%. (This assumes that the *S* would randomly distribute his errors on the verbal response between shock and nonshock syllables.)

While the Lazarus and McCleary experiment does not demonstrate that the threshold of a GSR response is lower than the threshold for verbal identification, nonetheless, their finding that in the presence of an error in the verbal response system the GSR response has better than chance accuracy has some important implications for a concept of behavior without awareness. Elsewhere the present author (1956a) has analyzed the subception experiment in terms of a multiple concurrent response model. In the Lazarus & McCleary experiment the *S* gave two concurrent responses to a stimulus—a verbal response and a GSR. The obtained subception effect consists of a partial correlation between the stimulus and GSR when the verbal response is held constant. Conceived in this way it is possible to specify the general circumstances under which such a partial correlation would be obtained.

The necessary and sufficient conditions are fulfilled when both responses are independently correlated with the stimulus but less than perfectly, and less than perfectly correlated with each other. The independent correlation of the two responses with the stimulus means that both the verbal response and the GSR are responses to the stimulus as perceived rather than the GSR being a response to the verbal classification of the stimulus. Less than a perfect correlation between the verbal response and the GSR can be alternatively stated as saying that these two response systems have noncorrelated error terms. Thus, variables which may affect the accuracy of use of verbal responses may not concurrently affect the accuracy of the GSR and vice versa.

Formulation of the results of the subception experiment in terms of partial correlation enables us to state specifically just what the subception experiment demonstrated. It succeeded in showing that both the verbal response and the GSR are each, at least to a degree, independently correlated with the stimulus and are partially independent of each other. In other words, the GSR is, at least partly, determined by the stimulus and the intervening perceptual process and not solely by the verbal response. This relation can best be seen in terms of

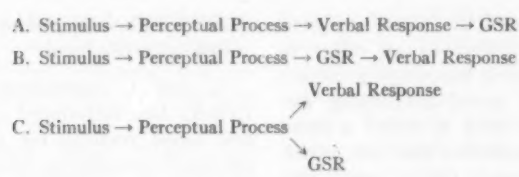
A. Stimulus → Perceptual Process → Verbal Response → GSR
B. Stimulus → Perceptual Process → GSR → Verbal Response
C. Stimulus → Perceptual Process
Verbal Response
GSR

FIG. 1.

the diagrams in Figure 1. If the relation was as depicted in A, then it would have been impossible to have obtained a subception effect. A partial correlation of the GSR with the stimulus holding constant the verbal response could not have been obtained. Similarly if the relationship was as depicted in B, it would be impossible to find a significant correlation between the verbal response and the stimulus when the GSR response was held constant. While the Lazarus and McCleary data are not presented so as to show whether or not a significant partial correlation between the verbal response and the stimulus existed when the GSR was held constant, in the replication of their experiment mentioned above a significant and appreciable correlation of this nature was obtained. This would seem to rule out the relationship as diagramed in B. This leaves only the relationship as diagramed in C where we have depicted the GSR and the verbal response as at least partially parallel response systems both independently responding to the perceptual process.

The author has carried out a series of experiments in his laboratory to test this partial correlation formulation and to see what implications it has for a concept of behavior without awareness. The subception effect has been successfully obtained in studying the concurrent GSR and verbal response generalization to a series of different size squares (Eriksen, 1956b). Here it was found that when the verbal response was held constant, a generalization gradient for the GSR as a function of square size was still obtained. In succeeding experiments verbal responses together with a concurrent lever movement response were studied (Eriksen, 1957, 1959). Again, if the instances where the verbal judgment or response was incorrect were analyzed the concurrent lever movement response showed better than chance accuracy. The effect also has been successfully obtained where the concurrent responses were an affective judgment of pleasant and unpleasant together with a specific verbal identification of tachistoscopically exposed stimuli (Eriksen, Azuma, & Hicks, 1959). In all of these studies a significant partial correlation was obtained between either response and the stimuli with the other response held constant. In no case, however, was it found that the nonverbal judgment gave a lower threshold or more accurate discrimination than did the verbal judgment.

These studies were also successful in indicating several variables that constitute the sources of noncorrelated errors between two concurrent response systems. Thus, it was found that distracting signals will introduce increased error into GSR responses without affecting the accuracy of concurrent verbal responses and similarly the number of competing alternative verbal responses affects the error of this judgment without concurrently influencing the accuracy of the GSR (Eriksen, 1956b).

On the basis of the above evidence the subception effect as an effect is real enough but it has nothing to say and offers no proof concerning the capacity of the GSR to reflect a more sensitive discrimination than that which is available through the *S*s' verbalizations. Instead the subception effect focuses attention upon the limitations of verbal responses in conveying the individual's perceptual experiences. By formulating the effect in terms of partial correlation in a concurrent response system model, we have the advantage of being able to direct our attention upon sources of noncorrelated errors between two or more concurrent response systems. We have already indicated some of the factors that will lead to increase in noncorrelated errors between concurrent responses and it is interesting to speculate on the effect that variables such as differential reward and punishment might have upon decreasing the correlation between verbal and certain nonverbal responses.

To say that the subception effect does not provide evidence for discrimination at levels below awareness is not to deny that the phenomenon is important to our general concept of awareness. There is reason to suspect that a considerable amount of the evidence that the clinician finds for unconscious processes is in the nature of discrepancy between concurrent responses. The clinician may note the *S*s' autonomic responses and evidences of emotion do not correspond, i.e., do not correlate very well, with his verbal statements. This lack of correlation is a basis for the clinician's inference of unconscious processes or defensive mechanisms at work. A concurrent response system model would seem the efficient way of conceptualizing this situation and suggesting ways of experimentally investigating the clinical phenomena.

Indirect Effects of Subliminal Stimulation upon Behavior

A recent experiment by Dixon (1958) has introduced an interesting and novel method for demonstrating effects of stimulation below the awareness threshold. Dixon determined the brightness threshold for a vertical and a horizontal line by the ascending method of limits. Having obtained a threshold Dixon told his *S*s that he was going to expose words to them at a brightness level where they would not recognize the words. He then exposed 10 words or phrases (6 were emotional in connotation and 4 neutral) at a brightness level .3 log ft. lamberts below the threshold obtained for the vertical and horizontal lines. The *S*s were told to give a word guess to each exposure even though they would not see the word. Each of the 10 words or phrases was exposed four times during a session and then at a later session the *S*s were presented the list of stimulus words and their responses were read to them. They were instructed to try to match their responses with the stimulus words using any type of associative connection they wished. Dixon's hypothesis was that the words exposed below threshold would have been unconsciously perceived in terms of their meaning and would thus influence the associations or responses the *S* made. Five of the seven *S*s gave results in this predicted direction. In fact, when the individual probabilities for each *S* were summed through *S*s the results were significant at a high level.

Dixon's approach to the study of discrimination without awareness is an interesting departure in methodology but subsequent work indicates his results are artifactual. Fuhrer and Eriksen (1960) in an extensive replication of the experiment confirmed that *S*s could match associations given to stimuli

that had been exposed at illumination levels too weak for correct identification. However, systematic employment of controls showed this effect to be a result of structural cues such as word or phrase length rather than "unconscious perception of stimulus meaning."

Fuhrer and Eriksen were unable to obtain significant or suggestive results at brightness levels of stimulus exposure below the threshold for vertical and horizontal lines obtained by the ascending method of limits or at the threshold obtained by a modification of Blackwell's forced-choice technique. However, when exposure was at a brightness level where *S*s could detect the presence of a stimulus but not identify it, significantly better than chance matching occurred. That the matching by these *S*s was due to structural cues rather than the perception of meaning was shown by the data from a control group. Using the same exposure level but now exposing the stimulus words and phrases upside down and backwards, the control *S*s were also found to give significantly better than chance matching of their associations to stimulus words. There was no significant difference in performance between this control group and the previous experimental group which had used the same exposure level. Since it seems extremely implausible that unconscious discrimination is capable of perceiving the meaning of words that are flashed upside down and backwards at very low illumination levels, it would appear that the performance of both groups is attributable to the cues perceived concerning the length of words or phrases.

But to check upon this possibility a further control was employed. For the experimental group a control list of words similar in structure but different in meaning to the actual stimulus words was constructed. Thus, for the stimulus phrase "male sex organs," the structurally similar control phrase was "made six organs." In the second session, when the experimental *S*s were asked to match their associations to the stimuli, they were also presented with this control list of structurally similar words together with the experimental words. Analysis of the results revealed no tendency for this group to match their associations with the correct words over the structurally similar control words.

The experimental work so far offers no convincing evidence that the human observer can respond differentially to stimulation that is of an intensity level below that at which a discriminated verbal report can be obtained. In other words, when adequate controls are provided, no response has been found to yield better absolute discrimination than a verbal response. However, there are most likely circumstances where a nonverbal response may be a better indicator than verbalization. No psychophysicist, for example, would expect an experimental arrangement where the *S* pressed a button directly in front of him to indicate "No" but had to get up and walk across the room to press a button to indicate "Yes" to yield as low a threshold as a situation in which both buttons were directly in front of the *S*.

Such response biasing can occur in less obvious ways. It is quite possible that anxiety conditioned to certain responses may bias their occurrence. Similarly, artificial restrictions on the number of verbal labels available can have a similar effect as well as deliberate response suppression due to social taboos associated with the response. But it is to be noted that under these circumstances one would not conclude evidence for a super unconscious dis-

crimination of stimulus intensities too weak to elicit verbal discrimination.

AWARENESS AND THE SUBJECT'S SUBJECTIVE CONFIDENCE

There is one other experimental approach to the problem of awareness that deserves brief consideration. It differs from the other methods in that a discriminated verbal report is not used as the criterion or definition of awareness. Instead awareness is defined in terms of the *S*'s statement as to whether or not he feels he has perceived the stimulus or whether he feels he is merely guessing. The evidence consists of showing that when the *S* expresses zero confidence, e.g., subjectively pure guessing, he nonetheless responds correctly to the presence or absence of the stimulus at better than chance level.

Peirce and Jastrow (1885) were among the first to report this phenomenon. Working with the difference limen on weight judgments, Jastrow's data indicate that in the presence of a subjective report of "no confidence," his judgments of heavier or lighter were still correct significantly better than chance. Since then this phenomenon has been a frequent finding in a number of experiments. In fact, as Adams (1957) points out, it is a stable enough phenomenon to permit its use as a classroom exercise. Representative of experiments obtaining results on this aspect of awareness are those of Fullerton and Cattell (1892), Miller (1939), Coover (1917), and Williams (1938).

Although these experiments are frequently and uncritically used to support contentions concerning discrimination without awareness where awareness is defined in terms of a discriminated verbal report, it is to be noted that these experiments do not fit this definition. Here a discriminated verbal report is obtained. The evidence for unawareness is in terms of the *S*'s report of lack of confidence in his verbal judgments. While this subjective confidence in his reports is an interesting problem, it has no operational similarity to the type of experiments we have considered above. An equation of awareness with verbalization, by definition, means that any verbal judgment is one made with awareness irrespective of whether the *S* has little or no confidence in the judgment.

The major difficulty with experiments of this kind is that they place upon the individual *S* the responsibility for establishing the criterion of awareness. We have no way of objectively knowing when *S* A says a given judgment was a pure guess and *S* B makes a judgment and states it was a pure guess that the two *S*s are talking about the same phenomena or that they are using the same criteria for guessing.

In work in our laboratory we have found very marked individual differences for the intensity range between where an *S* shows better than chance guessing in his responses and the point where he begins to report confidence in his judgments. For some *S*s the stimulus intensity where responses are better than chance almost coincides with their confidence threshold. For other *S*s there is a great range between where guesses are above the chance level and the point where confidence is reported in these judgments. While research on this particular aspect of awareness may be indeed fruitful, nonetheless, at the present stage of our knowledge we are not justified in lumping together data from these experiments with those that have been discussed in the preceding sections. Not only are they operationally dissimilar but more importantly, these marked individual differences indicate that as long as we rely upon the *S* himself to report whether or not a

verbal judgment was made with awareness, we are in fact using as many different criteria of awareness as we have experimental *S*s.

BEHAVIOR WITHOUT AWARENESS TO ABOVE THRESHOLD CUES

While we have been unable to find evidence for a supersensitive discriminating unconscious, the evidence that behavior can be directed by above threshold cues of which the *S* is unaware, is not only more plausible but somewhat more substantial. Common sense tells us that we are constantly utilizing cues of which we are unaware in our perception of depth, of shape and size constancy, as well as a number of cues that guide such complex motor habits as piano playing and driving automobiles. One of the factors making the elimination of such habits as fingernail biting or other mannerisms hard to eliminate is that they tend to occur without our awareness. Experiments appear to have demonstrated not only the use of cues without awareness but also the lack of awareness in some instances for the response the *S* is making, for the reinforcements controlling the response and in some cases the contingent relationship between cues, responses, and reinforcements.

Many of the experiments providing evidence on the effects of cues without awareness were considered by the *E*s as demonstrating effects of subliminal stimulation upon behavior. As will be seen the use of subliminal in connection with these cues was gratuitous since the particular experimental methodology did not contain the necessary controls or operations to permit the classification of these cues as subliminal. In fact, on the basis of other evidence there is every reason to suspect that these cues, when effective, were above a chance discrimination level for verbal report when assessed by proper psychophysical techniques.

The methodology employed in experiments in this area makes use of one or another of a few basic techniques. In most cases the cue is presented without forwarning the *S* that it will be present. Typically it is presented at an intensity so low as not to command attention. In other experimental arrangements the cue may be of a more obvious nature but the *S* is prevented from attending it by misdirection as to the purpose of the experiment or misdirection concerning the nature of the experimental task. In fact, misdirection of attention seems to be a crucial consideration in experiments of this nature, suggesting that perhaps we are dealing more with a concept of attention than one of awareness, if indeed the two concepts can be distinguished.

One may distinguish two separate lines of investigation using incidental cues. The first is concerned with the effect of incidental cues upon behavior that was learned outside of the experimental situation. For example, the effect of the incidental stimulation of the word "angry" upon the interpretation of ambiguous profiles might be studied.[2] The other approach is concerned with the problem of learning without awareness. The first approach we will term performance without awareness to distinguish it from the problem of learning without awareness.

Besides sharing the common use of the technique of incidental stimulation both lines of research investigation also share a common problem in the definition of awareness. Both must rely upon the interrogation of the *S* concerning his awareness in the experimental situation. As was pointed out

[2] I am indebted to George S. Klein for the term "incidental stimulation."

earlier in this paper such interrogation is fraught with difficulties. The value of the concept of awareness depends upon the adequacy with which the *E* surmounts these problems. The operational meaning of awareness will vary markedly from one experiment to the next depending upon the adequacy of the *E*'s assessment of the *S*s' verbalizations and the precautions taken to insure that they are valid and meaningful.

Performance without Awareness

There have been several experiments demonstrating that unaware cues are capable of producing perceptual illusions. Of the series of studies utilizing the Mueller-Lyer Illusion (Dunlap, 1900; Titchner & Pyle, 1907), the one by Bressler (1931) appears to be the most thorough and best controlled. He found evidence of illusory effects in the judgment of the lengths of the two lines when the directional arrows were so faint that the *S*s reported that they were unaware of their presence.

Miller (1939) and Perky (1910) projected geometric forms on the back of a ground glass screen at low intensities and found that the patterns presented were influential in determining the *S*s' choice of geometric pattern. The effect was found even when the *S*s were unaware that a pattern was actually being projected on the screen.

More recently investigators have made use of the metacontrast (Alpern, 1952) phenomenon as a method of presenting cues in such a way that the *S* would be unaware of them. Smith and Henriksson (1955) presented a stimulus consisting of fan shape lines, very briefly in a tachistoscope, and then, almost immediately after, flashed a square stimulus. While their *S*s did not report seeing the fan shaped lines, they did tend to see the square as trapezoid. Kolers (1957) has presented tachistoscopically sets of designs that required abstraction or conceptualization and found that if cues to the solution of the problems were flashed first, these cues were capable of influencing the type and direction of the solution even though *S*s reported they were unaware that the cues had been presented.

Using a modification of this metacontrast technique, Smith, Spence, and Klein (1959) presented a line drawing of an expressionless human face to their *S*s. The face was continuously presented to the *S*s except for brief interruptions where the word happy or angry was briefly flashed. They report suggestive evidence that the *S*s' interpretations of the face were more pleasant following exposures of the word happy than following exposures of the word angry (12 of 20 *S*s mentioned more pleasant characteristics of the face following exposures of the word happy than they did following exposures of the word angry). Exposures of the word angry however, had little or no effect upon the *S*s' interpretations of the face.

It is to be noted that the experiments described above in this section have been interpreted as demonstrating the effects of *subliminal* cues upon perception. If by subliminal we mean that the intensity of the cues was too weak to elicit a verbal report threshold under careful psychophysical conditions, then none of the above experiments permit interpretation in terms of subliminal stimulation. In Bressler's experiment even the faintest lines he employed, when studied independently of the illusion, were found to yield better than chance discrimination by his *S*s. Vinacke's (1942) data clearly show that the Miller and Perky results are not obtainable when the geometric forms are presented at an intensity level below a rather crudely determined

psychophysical threshold. Similarly the metacontrast effect seems to depend upon the first stimulus being at a level that would have been detectable were it not "washed out" by the second stimulation. And in the Smith et al. study, observations upon a control group of *S*s who had been alerted to the fact that words were being flashed indicated that most of the word exposures for the experimental group were above the level for detectability of the word. No evidence is presented to show that the presentation of the words happy and angry had a significant effect upon the interpretation of the profile when presented at durations below the detectability level.

Even when the above studies are considered as demonstrations of incidental rather than subliminal stimulation the conclusions must be accepted with caution. In most cases, the effects are slight even though statistically reliable. Careful replication of these studies is needed before confidence can be placed in the effect. And then, assuming replication substantiates the results, careful parametric analyses should be undertaken to explore not only the parameters of the experimental situation that create conditions where the *S* is inattentive to or unaware of the incidental stimulus but also studies oriented to studying the complexity of cues that can be utilized in this manner. Also, the effects of this incidental stimulation should be compared with the effects when the stimulation is actually detected in awareness.

There is another group of studies that would seem to fall under the classification of incidental stimulation but because of their particular nature deserve special consideration. These are the experiments based upon the Poetzl phenomena. In 1917 Poetzl reported a series of studies on his patients in which he exposed complex pictures tachistoscopically and instructed his patients to dream about the picture. He found that items in the picture which were not reported in intentional recall tended to occur in the dreams. Interest in this phenomenon was revived briefly in the early 1930's (Malamud & Linder, 1931) and then lay dormant until recently revived by Fisher (1954). An experimental report by Shevrin and Luborsky (1958) illustrates the methodology of research on this phenomenon as well as the general findings and methodological problems encountered. These investigators tachistoscopically presented a complex scene to their *S*s following which they asked for a description and drawing of what the *S* had seen. *S*s were then asked to bring in their dreams in a subsequent session. In this subsequent session further intentional recalls were obtained and the objects obtained in intentional recall were compared with the objects that appeared in the *S*s' dreams. Since many of the *S*s did not report dreams, they were asked to conjure up a mental picture or image which they were then asked to draw and to describe.

Shevrin and Luborsky report that a significant number of items and characteristics of the tachistoscopically exposed scene that were not mentioned in the intentional recalls occurred in the dreams or the *S*'s images. These findings were interpreted in terms of the psychoanalytical concept of day residues.

The complexity of these studies and experimental arrangements poses many methodological problems and gives rise to the possibility of numerous artifacts. One of the first problems encountered is the frequency or base rate with which different objects or items spontaneously occur in dreams and images independent of a prior pictorial stimulation. This is a particularly acute problem

since these investigators used a rather loose criterion for determining whether objects that had appeared in the picture occurred in the dreams or images. Thus, if the picture contained trucks and the *S* in his dream or image mentioned a bicycle, this was scored as an additional memory recovery since both represented a means of conveyance.

A second problem involved in the interpretation of Poetzl phenomenon data concerns the previously mentioned problem of the *S*s' subjective confidence level. When instructed to recall or describe what had been seen in the tachistoscopically presented picture we may assume that the *S* has a certain veridical orientation. That is, he wishes to be correct. As we have previously seen, *S*s have different criteria for reporting what they believe to have existed. Some *S*s may not mention an item unless they have fairly high confidence that it was present while others may use a lower confidence level. However, when freed of the veridical orientation required by intentional recall and asked to image, most *S*s would be expected to have freer criteria for what they would report. Thus, items which they were not certain were in the picture or for which they had low confidence are more apt to appear in the image than under instructions for intentional recall.

A third problem is what we might call the contingent probability of items going together in our enviornment. To illustrate what I mean by this, let us consider an *S* who has no recall of having seen a boat or a wharf in a picture but he does describe a body of water in his deliberate recall. When he is now asked to image, it seems likely that his associations are apt to be conditioned by the thought of water. Thinking of water it is likely that associations might run toward boats and wharfs and other objects that have in his past experience tended to be associated with a body of water. Since these objects would also have a fair probability of having actually been in the picture, this would lead to an increase in the recalled items in the image condition which was spurious due to the past learned associations among objects in our environment. In subsequent work on this problem a satisfactory control for this possibility might be obtained by having control *S*s who do imaging after observing the experimental *S*s' drawings and descriptions of the picture. Another control possibility would involve the use of stimulus pictures composed of heterogeneous objects whose inclusion was determined by random numbers.

Learning without Awareness

On the basis of the large number of studies during the past few years on learning without awareness, it would seem that the phenomenon itself should be well established by now. No attempt will be made here to survey the large literature. This has been well done by others (Adams, 1957; Krasner, 1958). Rather, this research area is mentioned in order to show that the phenomenon is not as well established as is generally supposed and also to indicate some of the directions that future research in this area must take.

As was also the case with research on performance without awareness, learning without awareness studies have been plagued with the problem of adequate operational criteria for awareness. Many of the experiments in this area have been concerned primarily with operant conditioning and little or no concern has been given to the definition of awareness. In many cases the only criterion is the statement by the *E* that the *S*s were un-

aware or that the *S*s were unable to label the correct response class. A few studies have focused primarily upon the phenomenon of awareness and here the evidence for learning without awareness has not been impressive (Krieckhaus & Eriksen, 1960). Adams (1957) in his survey has pointed out many of the inadequacies in this field of research such as the lack of replication of experiments, and more particularly the failure of *E*s to adequately rule out the possibility of correlated hypotheses.

More recently Dulany (in press) has called into question many of the verbal conditioning studies by showing that the results could be attributable to correlated hypotheses and the transfer of previously acquired verbal habits. Dulany used what has become a typical verbal conditioning procedure, studying the influence of the *E*'s "mmm hmmm" upon the *S*'s plural noun output. He found a significant increase in plural noun output by his *S*s although none of them were able to verbalize the correct response category. In this sense, then, they were unaware. But Dulany analyzed the data from his *S*s further. He found that the only *S*s who showed a learning effect were those who were able to verbalize to the effect that the *E*'s "mmm hmmm" indicated to them that they were associating in the proper category. He further demonstrated that associations to plural nouns tend to be other plural nouns whereas associations to singulars do not show this tendency. Since the *S*s were interpreting the "mmm hmmm" as an indication they were to continue associating in that category and the "mmm hmmm" followed a plural, this effect would be sufficient to account for the learning shown by his *S*s.

Research on learning without awareness as well as unconscious phenomena in general has been characterized by uncritical acceptance. Diven's (1937) experiment on unconscious conditioning of anxiety has been extensively cited and used as the basis for considerable theorizing concerning psychodynamics. Yet as Lacey and Smith (1954) have so clearly pointed out the Diven experiment did not contain the basic data comparisons necessary for a logical conclusion of conditioning without awareness. Using a more sophisticated design, Lacey and Smith, however, reported that they obtained evidence of unconscious conditioning of anxiety. But here again, Chatterjee and Eriksen (1960) have shown that Lacey and Smith's evidence on conditioning may well have been due to an artifact arising from the method of computing conditioning scores. These latter investigators were in general successful in replicating the Lacey and Smith findings but due to a more careful assessment of the *S*s' verbalizable awareness, they were able to show that the autonomic conditioning was no more precise nor specific than the *S*s' verbalizations.

Perhaps our tendency to uncritically accept experiments on unconscious phenomena may be due to our firm belief in the existence of the unconscious. But it is to be noted that while we might firmly believe in the phenomena of classical conditioning this does not discharge an *E* from the duty of showing that his particular conditioning results are not attributable to the artifact of "pseudo conditioning." There would seem to be little doubt that a considerable amount of human behavior occurs without awareness of the behavior at the time of its occurrence but it is to be noted that this does not logically require that behavior is learned without awareness. It may well be that the awareness process is biologically

adapted to be sensitive to features in the environment that are important to the organism and that any circumstances important enough to the organism to produce learning would not go undetected by awareness. This is not the same as saying that awareness is necessary for learning. Rather, it suggests that any situation where the cues and the reinforcements are salient enough to produce learning, will not escape detection by awareness.

It would seem that on the basis of the available evidence learning without awareness in human *S*s is not adequately proven. A considerable research effort is needed not only to arrive at acceptable operational criteria of awareness but to thoroughly explore the different types of learning situations to determine what if any kind of learning can take place without awareness. It may be that only simple conditioned responses are capable of learning without awareness or it may be that fairly high level concepts can be acquired. More attention in research needs to be paid to the awareness process itself, particularly in terms of criteria and independent correlates of awareness. Also, systematic parametric analyses of the experimental situation and arrangements where learning without awareness occurs must be undertaken. These latter studies should go far in suggesting an improved concept of awareness or attention.

SUMMARY AND CONCLUSIONS

This paper has surveyed the experimental work on discrimination, performance, and learning without awareness. The methodology employed in these experiments has been critically examined and the nature of the evidence evaluated. Some of the problems inherent in the definition of awareness have been considered in relation to the experimental work and an attempt has been made to show that there are a number of separate questions that have traditionally been lumped together under "unconscious" phenomena. Artifacts involved in experiments on subliminal perception and subception were pointed out and alternative interpretations for these experiments were suggested.

It was concluded that at present there is no convincing evidence that the human organism can discriminate or differentially respond to external stimuli that are at an intensity level too low to elicit a discriminated verbal report. In other words, a verbal report is as sensitive an indicator of perception as any other response that has been studied.

Consideration of the clinical evidence for the concept of the "unconscious" suggests that an assumption of "subliminal discrimination" is not required to account for the observed phenomena upon which the concept is based. Instead, the evidence noted by the clinician or student of personality would seem to be explainable in terms of either the demonstrable partial independence of different response systems to fragmentary perceptual cues and memory traces and/or to lack of awareness to above threshold cues.

A science of personality is not furthered by the frequent tendency of psychologists to discuss the "unconscious" with all the ambiguity and reverence that religions accord to the soul. There is great need to spell out explicitly the assumed characteristics of the unconscious and to search for explanations of so called unconscious phenomena in terms of more commonplace psychological variables. To do so may destroy the titillating mystery that the unconscious seems to hold but then that is the business of science.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

ALPERN, M. Metacontrast: Historical introduction. *Amer. J. Optom.*, 1952, **29**, 631–646.

BAKER, L. E. The pupillary response conditioned to subliminal auditory stimuli. *Psychol. Monogr.*, 1938, **50**(3, Whole No. 223).

BINDER, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62**, 119–129.

BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Engng. res. Bull., U. Mich.*, 1953, No. 36.

BRESSLER, J. Illusion in the case of subliminal visual stimulation. *J. gen. Psychol.*, 1931, **5**, 244–250.

CASON, H., & KATCHER, N. An attempt to condition breathing and eyelid responses to a subliminal electric stimulus. *J. exp. Psychol.*, 1933, **16**, 831–842.

CHATTERJEE, B. B., & ERIKSEN, C. W. Conditioning and generalization of GSR as a function of awareness. *J. abnorm. soc. Psychol.*, 1960, **60**, 396–403.

COOVER, J. E. Experiments in psychical research. *Psychical Res. Monogr.*, 1917, No. 1.

DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, **40**, 107–120.

DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *J. Psychol.*, 1937, **3**, 291–308.

DIXON, N. F. The effect of subliminal stimulation upon autonomic and verbal behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 29–36.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

DULANY, D. E. Hypotheses and habits in verbal "operant conditioning." *Psychol. Rev.*, in press.

DULANY, D. E., & ERIKSEN, C. W. Accuracy of brightness discrimination as measured by concurrent verbal responses and GSRs. *J. abnorm. soc. Psychol.*, 1959, **59**, 418–423.

DUNLAP, K. The effect of imperceptible shadows on the judgment of distance. *Psychol. Rev.*, 1900, **7**, 435–453.

ERIKSEN, C. W. Subception: Fact or artifact? *Psychol. Rev.*, 1956, **63**, 74–80. (a)

ERIKSEN, C. W. An experimental analysis of subception. *Amer. J. Psychol.*, 1956, **69**, 625–634. (b)

ERIKSEN, C. W. Prediction from and interaction among multiple concurrent discriminative responses. *J. exp. Psychol.*, 1957, **53**, 353–359.

ERIKSEN, C. W. Discrimination measured by multiple concurrent responses. *Psychol. Rep.*, 1959, **5**, 741–750.

ERIKSEN, C. W., AZUMA, H., & HICKS, R. B. Verbal discrimination of pleasant and unpleasant stimuli prior to specific identification. *J. abnorm. soc. Psychol.*, 1959, **59**, 114–119.

FISHER, C. Dreams and perception: The role of preconscious and primary modes of perception in dream formation. *J. Amer. Psychoanal. Ass.*, 1954, **2**, 389–445.

FREUD, S. *Collected Papers.* Vol. IV. London: Hogarth, 1925.

FUHRER, M. J., & ERIKSEN, C. W. The unconscious perception of the meaning of verbal stimuli. *J. abnorm. soc. Psychol.*, 1960, in press.

FULLERTON, G. S., & CATTELL, J. McK. On the perception of small differences. *Phil. Ser.*, 1892, No. 2.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychol. Rev.*, 1956, **63**, 149–159.

GOLDIAMOND, I. Indicators of perception: I. Subliminal perceptions, subception unconscious perception: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–412.

HILGARD, E. R., DUTTON, C. E., & HELMICK, J. S. Attempted pupillary conditioning at four stimulus intervals. *J. exp. Psychol.*, 1949, **39**, 683–689.

HILGARD, E. R., MILLER, J., & OHLSON, J. A. Three attempts to secure pupillary conditioning to auditory stimuli near the absolute threshold. *J. exp. Psychol.*, 1941, **29**, 89–103.

KOLERS, P. A. Subliminal stimulation in problem-solving. *Amer. J. Psychol.*, 1957, **70**, 437–441.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

KRIECKHAUS, E. E., & ERIKSEN, C. W. A study of awareness and its effects on learning and generalization. *J. Pers.*, 1960, in press.

LACEY, J. I., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

LAZARUS, R. S. Subception: Fact or artifact? A reply to Eriksen. *Psychol. Rev.*, 1956, **63**, 343–347.

LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: A study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.

MALAMUD, W., & LINDER, F. E. Dreams and their relationship to recent impressions. *Arch. Neurol. Psychiat.*, 1931, **25**, 1081–1099.

METZNER, C. A., & BAKER, L. E. The pupillary response conditioned to subliminal auditory stimuli: A control experiment. *Psychol. Bull.*, 1939, **36**, 625. (Abstract)

MILLER, J. G. Discrimination without awareness. *Amer. J. Psychol.*, 1939, **52**, 562–578.

NEWHALL, S. M., & SEARS, R. R. Conditioning finger retraction to visual stimuli near the absolute threshold. *Comp. psychol. Monogr.*, 1933, **9**, No. 43.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation:* Univer. Nebraska Press, 1957.

PEIRCE, C. S., & JASTROW, J. On small differences of sensation. *Mem. Nat. Acad. Sci.*, 1885, **3**, 73–83.

PERKY, C. W. An experimental study of imagination. *Amer. J. Psychol.*, 1910, **21**, 422–452.

POETZL, O. Experimentell erregte Traumbilder in ihren Bezichungun zum indirekten. *Sehen. Z. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.

SHEVRIN, H., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomena. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

SMITH, G. J. W., & HENRIKSSON, M. The effect on an established percept of a perceptual process beyond awareness. *Acta psychol., Amsterdam*, 1955, **11**, 346–355.

SMITH, G. J. W., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.

STECKLE, L. C. Two additional attempts to condition the pupillary reflex. *J. gen. Psychol.*, 1936, **15**, 369–377.

STECKLE, L. C., & RENSHAW, S. An investigation of the conditioned iridic reflex. *J. gen. Psychol.*, 1934, **11**, 3–23.

STERN, F. An investigation of pupillary conditioning. Unpublished doctoral dissertation, Univer. Washington, 1948.

TAYLOR, F. W. R. The discrimination of subliminal visual stimuli. *Canad. J. Psychol.*, 1953, **7**, 12–20.

TITCHNER, E. B., & PYLE, W. H. The effect of imperceptible shadows on the judgment of distance. *Proc. Amer. Phil. Soc.*, 1907, **46**, 94–109.

VINACKE, W. E. Discrimination of color and form at levels of illumination below conscious awareness. *Arch. Psychol.*, 1942, **38**, 15–16.

WEDELL, C. H., TAYLOR, F. V., & SKOLNICK, A. An attempt to condition the pupillary response. *J. exp. Psychol.*, 1940, **27**, 517–531.

WILCOTT, R. C. A search for subthreshold conditioning at four different auditory frequencies. *J. exp. Psychol.*, 1953, **46**, 271–277.

WILLIAMS, A. C., JR. Perception of subliminal visual stimuli. *J. Psychol.*, 1938, **6**, 187–199.

(Received January 8, 1960)

Psychological Review
1960, Vol. 67, No. 5, 301–316

SOME IMPLICATIONS OF A SOCIAL LEARNING THEORY FOR THE PREDICTION OF GOAL DIRECTED BEHAVIOR FROM TESTING PROCEDURES[1]

JULIAN B. ROTTER

Ohio State University

Many sophisticated observers are aware that a wide gap exists between personality theory and the techniques or procedures used to measure personality variables. The low level of prediction of such testing procedures may well be a function of the failure to apply the theories themselves to the methods of measurement. Particularly, it is a failure to apply an analysis of the determinants of behavior in general to the specific test taking behavior of the subjects (*S*s).

The gap itself may be described as having three aspects. The first of these relates to the question of the constructs used in the theory and the constructs which the tests were developed to measure. In many instances rather than devising tests which measure specific theoretical constructs which are carefully defined and for which the test behavior can be understood as a logical referent, the descriptive constructs used to classify test response do not logically relate to the new theoretical constructs but are bent or twisted to measure the new variables. That is, test constructs which were used to classify test responses developed earlier are "translated" to be measures of the new variables. Examples of this are use of Rorschach variables such as color, movement, and shading which arose from imagery-type theory, to measure such constructs as "ego strength," "rigidity," and "tolerance for ambiguity." The Rorschach was not developed to assess such variables and in translating or twisting older methods of Rorschach scoring to measure these variables it is quite likely that a loss of prediction results.

A second aspect of this gap between personality theory and methods of measurement of personality lies in the testing procedure itself. For example, where the theory may emphasize the significance of differences in behavior in the presence of authority figures vs. peers or males vs. females, the formal test procedure assumes no such variables are important. That is, no difference in interpretation of test results follows from the fact that the examiner may have a different social stimulus value in one case than in another or under one set of conditions rather than another. In such an instance although the theory itself recognizes (and experimental data such as Gibby, Miller, & Walker, 1953, and Lord, 1950, support) the importance of the effect on behavior of the nature of the social stimulus, the test procedure itself does not take it into account. An example would be in the application of Murray's theory (1952) which sees behavior as a function of internal *needs* and environmental *presses*. Tests have been developed using this theory (Thematic Apperception Test, as clinically used; Edwards Personal Preference Schedule) which presume to measure the strength of various needs but fail to account for the test behavior as a function of the testing situation itself (an

[1] I am indebted to Shephard Liverant for his helpful comments and suggestions about this paper.

environmental press) as one of the variables determining the test behavior. Other characteristics of this discrepancy between theory and test taking procedure will be discussed more fully later.

A third aspect of the gap lies in the area of inference from test behavior. The issue here is that there is an absence of logic or contradiction in the assumed relationship between what the *S* does, or test behavior, and what is inferred from such behavior. Peak (1953) and Butler (1954) among others have discussed this problem earlier. Jessor and Hammond (1957) have noted such a gap in the inferences made from the Taylor Anxiety Scale. Another example could be drawn from the Edwards Personal Preference Schedule (Edwards, 1953) in which *S*s are asked to state their preferences for different kinds of goals but there is no theoretical basis provided to allow one to make predictions about *nontest behavior* from such preferences. Of course, it can be assumed that the preferences have some one-to-one relationship with some criterion behavior, but it is unlikely that even the test authors would make such a theoretical commitment. In other words, it is not clear exactly what can be predicted or should be predicted from the test responses. Individuals using such tests, however, can defend themselves by stating that prediction is after all an empirical matter and one has to find out what can or should be predicted. It is likely, however, that the construction of tests which are systematically or theoretically pure, in that they are devised to measure specific variables or to make specific predictions, with the method of measurement and inference consistent with the theory will ultimately provide much better predictions of behavior as well as a test of the utility of the theory itself.

The purpose of this paper is to explicate some of the implications of a social learning theory of personality for the measurement of personality variables. The particular point of emphasis is the measurement of goal directed behavior conceptualized in social learning terms as *need potential*. Secondarily, the paper aims at illustrating the nature of the relationship between testing procedures and inference about behavior more generally.

In social learning theory (Rotter, 1954) the basic formula for the prediction of goal directed behavior is as given below:

$$BP_{x,s_1,R_a} = f(E_{x,R_a,s_1} \,\&\, RV_{a,s_1}) \quad [1]$$

The formula may be read as follows: The potential for behavior x to occur in Situation 1 in relation to Reinforcement a is a function of the expectancy of the occurrence of Reinforcement a following Behavior x in Situation 1 and the value of Reinforcement a, in Situation 1. Such a formula, however, is extremely limited in application for it deals only with the potential for a given behavior to occur in relationship to a single specific reinforcement. The prediction of responses from personality tests requires a more generalized concept of behavior and the formula for these broader concepts is given below:

$$BP_{(x-n),s_{(1-n)},R_{(a-n)}} = f(E_{(x-n),s_{(1-n)},R_{(a-n)}} \,\&\, RV_{(a-n),s_{(1-n)}}) \quad [2]$$

This may be read: The potentiality of the functionally related Behaviors x to n to occur in the specified Situations 1 to n in relation to potential Reinforcements a to n is a function of the expectancies of these behaviors leading to these reinforcements in these situations and the values of these reinforcements in these situations. For purposes of simplicity of communication, the three basic terms in this formula have been

typically referred to as need potential, freedom of movement, and need value as in the third formula below:

$$NP = f(FM \& NV) \qquad [3]$$

In this formula the fourth concept, that of the psychological situation, is implicit. The variables referred to above and operations for measurement have been defined and further explicated in a previous publication (Rotter, 1954).

In order to illustrate the social learning theory implications for measurement of personality and for the measurement of goal directed behavior, it seems expedient to consider three basic approaches to this problem based on the number of determinants used theoretically to predict such goal directed behavior and the problems, limitations, and advantages of each approach.

STRENGTH OF NEED AS A BASIS FOR PREDICTING BEHAVIOR

Although many esoteric systems of prediction utilize essentially the strength of need, drive, or instinct approach, this method can be described as the simplest or least complicated approach. Basically a series of constructs are formulated more often on an a priori basis than empirically, or at least on a presumed clinical rather than experimental basis. These descriptive terms may refer to instincts, drives, needs, factors, entities, or vectors of the mind (i.e., the Minnesota Multiphasic Personality Inventory, Edwards Personal Preference Schedule, Rorschach, Humm-Wadsworth, etc.). They all have in common that there is more than one basic characteristic, that these two or more characteristics are in some way measurable along a continuum and presumably the individual's behavior can be predicted from the characteristics which are "stronger" and the characteristics which are "weaker."

Sometimes the personality disposition can be predicted from the strength of other constructs according to either simple or complex statements of relationship formally postulated, hypothesized, or informally asserted. These relationships can become quite complex as in psychoanalysis or quite esoteric as in Szondi's explanation that motivated behavior is a result of the interaction of dominant and recessive genes. Because the methods of measurement in some instances cannot be direct, as in the assessment of unconscious drives in psychoanalysis, an impression of great complexity is given but regarded entirely from the point of view of the prediction of behavior, the system may still have a simple character. The potential for a given kind of behavior is still directly predictable from the strength of the drive, instincts, needs, or energies postulated.

There is another form of this model in which the various drives, dispositions, or needs are considered to interact. For example, the individual may be conceived of as being controlled by his intellect and his emotions, but his behavior must be understood in light of the interactions of these two forces with a third variable, the will as in the Rorschach Test (1942). Again, this makes complex the calculation of strength or weakness but does not change the overall method of making predictions. Whether dealing with will, intellect and emotions or ego, superego and id, the effect of interaction is only to increase or decrease the tendency of one of the needs to function or to strengthen or weaken one of them or perhaps to produce a fourth or fifth additional need. The basic method of prediction stays the same although the calculation of strength or weakness in such a model becomes more difficult.

The obvious problem, of course, for such rudimentary method of prediction

is how to predict anything at all. If a system included five instincts or needs and these are ordered on some metric system from high to low, does one assume that the person will always act in the fashion to be predicted from his strongest or highest need? If an *S* is more oral than anal, does he always act in an oral fashion? Actually the most logical assumption in regard to any specific instance is that he will always act the same way. One might presume on a statistical basis, as it is sometimes done, if the individual is at the 70th percentile on Need A and at the 30th percentile on Need B, 70% of the time he would act in one fashion and 30% of the time in the other over some undefined period of time. However, the only sensible statistical or logical prediction in any specific instance, if no other variables are concerned, is that he would act in accordance with the higher need. This might still give fairly good prediction if only 2 variables are involved, but if 20 variables are involved and many of them are very close in value or "strength," then the amount of error begins to increase. In fact, it becomes a problem to predict even slightly above chance and, indeed, except for some limited and highly controlled experimental situations, this is the problem in psychology now. A recent illustration of this failure is reported in a carefully controlled study by Little and Shneidman (1959) who failed to find much relationship between interpretations of psychological tests (Rorschach, MAPS, TAT, and MMPI) and anamnestic data. Loevinger (1959) summarizing the predictiveness of individual tests in the recent *Annual Review of Psychology* states, "To date the only tests which meet standards for individual prediction are those of general ability" (p. 305). Previous reviewers have made similar statements.

Another problem which arises in the prediction of behavior with this simple model is that it soon becomes apparent that the strength of a wish, need, or drive to achieve some goal such as being taken care of, obtaining love, or injuring someone is not a good predictor of the occurrence of behavior directed towards the achieving of that goal. To some extent this problem can be dealt with by the notion of interaction of needs, but usually in order to account for the discrepancy between need or wish and behavior, constructs of the same order do not provide sufficient explanatory basis. It is actually necessary to postulate some other kinds of internal constructs to account for the discrepancy between what might be called wish, desire, need, or instinct and observable behavior.

In the measurement of these need strengths all varieties of tests and devices have been used. To some extent, the personality questionnaire is utilized a little more by people adopting such a predictive scheme as that described above, but also projective tests, observation, interviewing, and many other techniques of personality measurement have been used in this fashion. Test construction methodologies may currently be more sophisticated in that they control for social desirability of items, motivation, faking, lying, and inability to understand directions. Recent tests may also rely on purification of factors, cross-validation, or item analysis. However, with all these "modern improvements" in test design one is still left with a series of figures which are of doubtful utility for the actual prediction of behavior at a level satisfactory for either practical application or for the clarification of theoretical issues.

THE ADDITION OF AN EXPECTANCY CONSTRUCT IN THE PREDICTION OF GOAL DIRECTED BEHAVIOR

The absence of additional variables explicitly defining the relationship of need and behavior appears to be not so much a matter of simple theoretical structure as it is merely the absence of any real explicit theory about human behavior. The development of a predictive model which recognizes the discrepancy between need and behavior and tries to systematically take it into account represents a second level of sophistication.

At an earlier date perhaps psychoanalysis dealt with this problem most effectively in introducing concepts such as repression, sublimation, suppression, defense, reaction formation, etc., to account for the discrepancy between observed behavior and the presumed internal drive, need, or instinctual urge.

At this level of theorizing some systematic variable is added to the internal motivational state in order to predict behavior. Perhaps another way of saying this is that in addition to some measure of preference or value of a specific goal another systematic concept must be introduced, not only in a hit-or-miss fashion but perhaps directly into our assessment procedure. The psychoanalytic solution has been criticized because many specific concepts are introduced to account for a discrepancy between drives, urges, or needs and observable behavior, but these concepts are not readily measurable. In addition, one does not know when one explanation, i.e., reaction formation, is the explanatory concept or another such as sublimation or simple repression.

In social learning theory (Rotter, 1954) it is presumed that the relationship between goal preference (reinforcement value) and behavior can be determined only by introducing the concept of the individual's expectancy, on the basis of past history, that the given behavior will actually lead to a satisfying outcome rather than to punishment, failure, or, more generally, to negative reinforcement. Since the early formulations of Tolman (1932), expectancy theories have become more and more widely relied upon both in human learning and personality theories. It is possible to conceptualize more specific constructs such as repression and reaction formation as only special cases of an expectancy for severe punishment and that a more general relationship holds which includes perhaps all of these and also expectancies for punishment or failure of which the individual is quite aware. For example, an individual may wish very much to be a good dancer and to dance with members of the opposite sex. He makes no attempts, however, to dance at a party or a dance because he can tell you "but I look like a fool when I go out on the dance floor." We need, in other words, to introduce no specific construct involving the "unconscious" to explain the discrepancy between his wish and his behavior. The *S* may or may not be aware of expectancies which influence his behavior. Whether or not he is aware of them may affect the degree to which these expectancies change with new experience as well as other variables. The degree of awareness may be an important additional variable; however, the level of expectancy itself is the broader variable which bears a direct relationship to the potential occurrence of a specified behavior.

The question arises, then, of how one takes into account such factors in an actual testing situation. It could be said that no one is really so naive as to believe that the strength of an internal motivational condition or need is a di-

rect predictor of behavior. Somehow or other, whether or not the individual had learned a given behavior or expected it to work is also an important aspect of prediction. However, more often than not this aspect of prediction has been treated as a source of error, something to be eliminated if possible, both in testing or in the validation of a test instrument. As a matter of fact, many currently used instruments attempting to assess the strength of motives, drives, or needs are usually confounded. Although they may be quite sophisticated in methodology, the test items or the test variables usually refer in part to what the individual did, i.e., overt behavior ("I frequently lose my temper"), in part to what he wished ("I would like to have more friends"), and in part to what he expected to be the outcome of his own behavior ("I feel that other people do not appreciate my good intentions"). To some extent these impure items probably add to prediction by providing more than one kind of referent for behavior, but the nonsystematic way in which they are used also limits prediction.

In trying to predict goal directed behavior from tests, two possibilities are open. One of these is to attempt to predict behavior from other behavior which presumably is functionally or predictively related to the test behavior. What this involves is analyzing test situations as behavioral samples under a given set of test conditions. For example, to assess dependent behavior one could use direct observation techniques, perhaps in problem solving situations, in which the *S* is scored for help-seeking behavior (cf. Naylor, 1955). In questionnaires the items should refer to what the *S* does, not to what he expects, wishes, or feels. The use of behavior samples for predictions or the regarding of all kinds of tests including projective tests essentially as samples of behavior to be analyzed in terms of what the *S* does under these conditions has been described elsewhere (Rotter, 1954). Like the work sample test in industry it undoubtedly provides the best prediction to a limited specific behavioral criterion since it requires the fewest intermediate constructs and the fewest assumptions regarding the action of other variables.

However, there are many problems, both theoretical and clinical, when it is important to break down this behavioral measure into its major determinants of reinforcement value and expectancy for the occurrence of the reinforcement. For example, in psychotherapy an understanding of how some behavior or group of behaviors may be most readily changed requires analysis into at least these two components. Even when strictly concerned with predictions of behavior in a broad band of life situations, rather than change, analysis into separate determinants may provide greater prediction than a work sample or behavioral technique. In this second alternative it is important either to control or systematically vary the other variable or measure both. For example, Liverant (1958) has measured some needs by presenting pairs of items involving goal preference matched for social desirability, and Jessor and Mandell (Mandell, 1959) are developing a similar test to measure expectancy for success in satisfying the same needs.

In using projective material such as the TAT, Crandall (1951) has demonstrated that expectancy for need satisfaction, for which the term freedom of movement is used in social learning theory, can be reliably measured by selecting particular kinds of referents. The work of Mussen and Naylor (1954), Kagan (1956), and Lesser (1957) gives strong evidence that the

relationship between theme counts of aggression on the TAT and overt aggressive behavior depends to a large extent on whether or not that overt behavior is socially acceptable in the *S*s' own homes or social climates. The relationship between theme count and overt aggressive behavior appears to hold only when the *S*s do not have a high expectancy that aggressive behavior will be punished. Atkinson and Reitman (1956) report that in a number of studies of need achievement, it has become clear that prediction of behavior is enhanced if, in addition to taking a measure of need achievement based upon achievement theme counts in TAT-like material, an additional measure of expectancy for success is also taken into account.

In dealing with this type of testing material the recently published study of Fitzgerald (1958) provides a more systematic analysis. Using a highly reliable sociometric technique of nomination of fraternity brothers as his behavioral criteria and dealing with the need dependency, Fitzgerald found no relationship between theme counts and overt behavior. Presumably, dependent behavior is not socially acceptable among male college students. He had, however, independent interview ratings of need value, that is preference or desire for dependency satisfactions and of freedom of movement, or expectancy that behavior directed toward achieving dependency would be satisfied.

He found that by using these measures he did obtain a significant correlation between theme counts and the *discrepancy* between need value and freedom of movement. More specifically, what he called a conflict score or score indicating the degree to which the individual preferred dependency or desired dependency satisfactions but expected that he could not achieve them correlated with theme counts for dependency.[2] On the other hand, an Incomplete Sentences Blank measure of dependency which utilized behavioral referents as well as reinforcement value and expectancy referents did show a low but significant relationship of the number of completions dealing with dependency with both the sociometric and interview measures of need potential or actual dependent behavior. Although an actual analysis was not made, it seems very likely that the reason for the correlation in the case of the ISB and not the TAT is that at least some of the ISB completions were descriptions of actual behavior. Possibly a purer measure of behavior would have shown a greater relationship to the sociometric and interview rating assessment of actual dependent behavior in life situations.

Should we build two instruments or at least two sets of testing operations to separately assess need value and freedom of movement, or should we attempt to use behavioral measures in order to make our predictions about behavior, we would still be faced with the problem of predicting in a specific situation. Given measures of six behavior potentials, however arrived at, the problem remains that of knowing which of these is likely to be the behavior preferred in some specific situation. One is again forced to predict that the behavior with the highest potential always occurs and one is limited again in prediction to a very low level of accuracy. In the laboratory situation where we can reduce the possible alternatives to two, significant, although not predictive results, are possible. In the

[2] Whether or not the relationship between theme count and high reinforcement value and low expectancy is general is not yet known. It appears at this time to possibly depend on whether or not the test material and testing situation is conducive to the free expression of fantasy.

life situation where the alternatives are very frequently of a large order, the question arises of whether or not any useful prediction is possible. This leads us to a third level of sophistication, one in which the psychological situation is one of the variables on which prediction is based.

The Psychological Situation as a Third Determinant of Goal Directed Behavior

Few would deny that the psychological situation will affect the potential of occurrence of any behavior or class of behaviors. However, the fact that behavior will vary from situation to situation is most often treated as a source of error, something to be avoided. If possible, one should construct tests or find personality variables which rise somehow above the situation. It is probably no exaggeration to say that thousands of hours of wasted work have been done by psychologists in the vain goal of finding either tests or variables which would, somehow or other, predict regardless of the situation in which the test is given or regardless of the situation in which the predicted behavior is expected to occur.

There are three separate problems here which will be discussed as one basic problem. The first problem is to understand the effect of the testing situation on test results. For example, Sarason (1950) has provided an excellent discussion of some of the influential situational variables in intelligence testing. The second problem is to understand the nature of the criterion situation which affects the criterion measures. The third and ultimately most important problem is to devise our tests with full consideration of the nature of the test situation in order to predict behavior in other situations for which the test was constructed. In other words, we need to devise tests not to predict personality or needs or behavior in the abstract but in specified situations or classes of situations if we want high prediction. *We need to know and take into account the dimensions of situation similarity in devising test procedures.*

Cronbach (1956) has criticized the failure to regard differentially the criterion situations in which tests are applied. In regard to the test situation we have only attempted to standardize the test procedure but usually have ignored the importance of the social context in which the test is given. Perhaps the most important thesis of this paper is that the psychological situation needs to be understood and systematically considered in our predictive formula, not treated as a source of error or something that can be ignored because part of the total situation is standardized.

Recently there have been a number of studies which demonstrate that almost all tests are subject to faking, to instructional variation, to examiner influences, to testing conditions, etc., regardless of the type of test (Borislow, 1958; Green, 1951; Gross, 1959; Mussen & Scodel, 1955; Rotter, 1955). The general inference drawn from these studies is that the tests are poor. Actually the implication of such findings is that we are making inefficient use of our tests. If the test situation for many personality tests is one in which social conformity or acceptability is easily achieved and no other satisfactions are given up in achieving acceptability, then for some purposes this motive should be controlled. However, the test situation can also be utilized to measure the importance of social conformity for the individual. *What we call faking is only our recognition of the fact that the S is taking the test with a different purpose or goal than the one the examiner wants him to have.* For

some purposes it might be important to understand what kind of goals he exhibits in this kind of situation. More often than not we simply try to control what we should be studying. For example, in giving intelligence tests it might be better to study systematically the effect on performance of encouragement and discouragement rather than to attempt some mythical neutral attitude which is presumably the same for every examiner. Knowledge of the effects of situational variations would be of particular value in understanding the frequently diverse and contradictory results of apparently similar research investigations. For example, Henry and Rotter (1956) found that large, predicted differences were obtained between two comparable groups on the Rorschach test if one group was reminded before the regular instructions that the test had been used frequently to study psychopathology. An obvious implication of this study is that investigators using this same test in the college laboratory and in the clinical setting may well produce diverse results.

Another example of how the situation can be used in testing is provided by the patient who is being assessed for possible benefit from psychotherapy. If the clinic or hospital can provide both male and female therapists and also therapists who rely on support and direction as well as therapists who remain distant and passive, then the testing procedures can be varied so that those situational influences are present. The testing could provide information to indicate what kind of therapy and what kind of therapist is likely to provide *this* patient with the most efficient conditions for relearning. For more conventional purposes of clinical testing, it is still more important to know under what conditions the patient behaves in a paranoid fashion and under what conditions he does not, than it is to know how many percentiles of paranoia he has.

Not only can the test situation itself be analyzed as a behavioral sample but situational referents can be incorporated into the content of items by systematic sampling. For example, questionnaire items can deal with the *conditions* under which the *S* feels tense, nervous, happy, has headaches, etc. as Mandler and Sarason (1952) have done for some intellectual test taking situations. Similarly, projective methods, particularly TAT-type tests, can systematically vary the situation through the selection of test stimuli as has been done by Crandall (1951) and McClelland, Atkinson, Clark, and Lowell (1953). More recently Murstein (1959) has suggested a conceptual model for stimulus variation with thematic techniques.

The many studies indicating marked effects of testing conditions suggest that it is of great importance in the publication of any test that descriptions of the differences in test results that are likely to be associated with different kinds of testing situations be provided. No test can be adequately understood unless the data regarding its standardization or use includes systematic descriptions of the differences in test results which are a function of different kinds of testing conditions and different kinds of purposes in taking the test for similar samples of *S*s. Only when we know whether an *S* is likely to produce different test results when he is taking the test to demonstrate how imaginative he is as compared to taking it to prove that he needs help will we be adequately able to understand the meaning of test results and to predict future behavior from them.

There have been personality theorists who have made much of the importance of the individual's life space. Kantor (1924) was one of the first to empha-

size that the basic datum of psychology is the interaction of an individual and his meaningful environment. For Kantor, people do not have internal characteristics in the same sense as for other theorists; rather they have a reactional biography of interactions with the environment. Lewin (1951) has also emphasized the importance of the life space or psychological situation in the determination of human behavior. Brunswik (1947) has repeatedly called for analyses of and sampling of psychological situations for predictive purposes. Helson (1948) has applied his theory of Adaptation Level to social psychology stating that the effect of the total field can be quantified by careful ordering of the field of exposed stimuli. Recent concern with the importance and need for systematic study of situation variation has been expressed by Allport (1958) and Cronbach (1957).

In a more limited and less systematic way, psychoanalysis has suggested, in a few areas, that certain kinds of goal directed behavior depended upon the psychological situation. This is done in making distinctions between the individual's potential response to authority figures vs. nonauthority figures and males vs. females. Beyond this, little systematic analysis of differences in life situations has been made by the traditional analyst. Murray's (1952) formulation of the nature of personality stresses that behavior is a function of the interaction of an individual with a psychological situation which he felt could be categorized as "press." At a more specific level Atkinson and Raphelson (1956) have shown the value of including situational variables in studying achievement behavior. This general point of view has also been represented in sociology by Thomas (1951) and Coutu (1949).

In social learning theory, it has been hypothesized that the situation operates primarily by providing cues for the *S* which are related to the magnitude of his expectancies for reinforcement for different behaviors. The effect on the value of the reinforcement itself operates through expectancies for associated or subsequent reinforcements which may differ from situation to situation. It has also been hypothesized that situations may be usefully categorized in terms of the predominant reinforcements as culturally determined for any large or small culture group. There are many other possible ways of categorizing situations depending upon the predictive purposes involved.[3] Methods of determining generality or determining the dimensions of similarity among situations have been described in an earlier paper (Rotter, 1955).

[3] Several writers have pointed out the difficulty of identifying situations independently of behavior. That is, how can one describe a situation as one would a physical stimulus independently of the *S*'s response? The problem is not different from that of describing stimuli along dimensions of color although perhaps vastly complicated in social or other complex situations. In the case of color stimuli ultimately the criteria is a response of the scientist or observer, sometimes a response to an intermediate instrument, and one that is at the level of sensory discrimination and so leads to high observer agreement. In the case of the social situation, the level of discrimination is common sense based on an understanding of the culture rather than the reading of an instrument. As such, reliability may be limited but still be sufficiently high to considerably increase prediction. In this way specific situations could be identified as school situations, employment situations, girl friend situations, etc. For the purpose of generality various kinds of psychological constructs could be devised to arrive at classes of situations which have similar meaning to the *S*. The utility of such classes would have to be empirically determined depending on the *S*'s response. The objective referents for these situations, which provide the basis for prediction, however, can be independent of the specific *S*. That is, they can be reliably identified by cultural, common sense terms.

Two illustrative studies of an increasing number of experimental analyses of behavior which vary both internal characteristics and the psychological situation systematically in the same study are described below. These studies follow the basic paradigm that the presumed relevant individual (personality) and situational (experimental manipulations) variables can be observed simultaneously and their interaction studied.

ILLUSTRATIVE STUDIES VARYING BOTH THE SITUATION AND INTERNAL CHARACTERISTICS

A recent doctoral dissertation by James (1957) illuminates very clearly the potential of greater prediction when both the situation and the internal characteristic are varied in the same study. The behavior being studied by James involved a variety of learning variables, including acquisition, changes or shifts, extinction, generalization, and recovery of verbalized expectancies for gratification. Two general hypotheses were involved in this study growing out of previous work by Lasko (1952), Phares (1957), Neff (1956), and James and Rotter (1958). Hunt and Schroder (1958) have also dealt with what appears to be a related variable. The first of these hypotheses might be stated as a situational one. That is, that the nature of a learning process differs depending upon whether or not the situation is one in which the reinforcements that occur are a direct outcome of some internal characteristic of the individual such as skill, a physical characteristic, or whatever, versus a situation where the reinforcements are essentially controlled by someone else or by chance or by conditions or powers beyond the *S*'s control. Perhaps a good example of the latter would be a dice game or the winning of a door prize or having soup spilled on one because a waiter tripped, etc. James utilized line and angle matching tasks reinforcing each *S* positively on his guesses in six of the eight training trials. He specifically hypothesized when the situation is structured in such a way that the *S* expects the occurrence of reinforcements to be beyond his control or partly beyond his control, increments and decrements in expectancy for gratification as a result of experience are smaller, the number of unusual shifts, that is, shifts up after failure or down after success, are greater, extinction is faster, and there is less generalization from one task to another and greater recovery following extinction.

The measurement of individual differences in this study followed from the previous work of Phares which suggested that individuals can be differentiated in the degree to which they see the world and the things that happen to them as controlled by others or as determined by chance or unpredictable forces. The second hypothesis, then, was that all the differences which would occur as a result of the situational conditions would also be true of individuals within all groups as a function of their general attitude towards "control of reinforcement."

In order to predict the individual differences in attitude, James enlarged and revised the questionnaire first devised by Phares. This was given to all *S*s at the end of each experiment. The results are most striking. All of the predicted outcomes hypothesized above resulting from the differences in instructions or situations were obtained and all were statistically significant with the exception of recovery following extinction, which showed a strong trend in the predicted direction. Similarly, within each group the individuals high as compared to low on the questionnaire differed significantly in ex-

actly the same way as did the groups themselves as a result of the different instructions or situations presented. Although individual prediction was not the concern of this investigation, it is quite clear that a simple formula could be devised which could predict all of the learning variables involved in this study with a fair degree of accuracy. Certainly it is clear that a greater degree of accuracy is possible when both the situational and individual variables are taken into account. Perhaps far more important, this study indicates that various experimental paradigms in studying human learning are likely to produce different kinds of results. Whether or not a given learning task is one in which the *S* feels that success is dependent upon the experimenter's manipulation (for example, when he is expecting to predict a random sequence of red or green lights) or is the result of his own skill provides a crucial difference in the nature of the learning process itself.

The study of James, however, does not provide a satisfied feeling that it illustrates all of the problems of prediction involving both the individual's characteristics and the situation from which the prediction is made. It gives an almost too simple picture of the interaction of these two variables. Another recent dissertation by Moss (1958) suggests that this relationship can be more complex, and illustrates more clearly the effect of the testing situation on more commonly used types of assessment procedures.

Moss studied a general behavioral characteristic which he called cautiousness. Essentially this was defined as the avoidance of risk, the selection of the safest alternative in a situation where failure or negative reinforcement was possible. He varied the situation by reacting differently to three groups following the administration of a questionnaire which he described as a measure of social acceptability. One group was shown false norms at the conclusion of the questionnaire that indicated that they were in the ninetieth percentile of social acceptability for a college group. Another group of *S*s was shown that they were at the tenth percentile, and a third group was given no information about the results of this supposed test of social acceptability. He hypothesized that cautious behavior would increase with negative reinforcement. Immediately following this procedure, the *S*s were given two projective type tests and behavior on these tests was analyzed as to degree of cautiousness.

Prior to the giving of the "social acceptability" questionnaire the *S*s had been tested on the level of aspiration board. Behavior on the level of aspiration board (Rotter, 1942) was categorized into cautious or noncautious patterns.[4] The general tendency to seek safe alternatives in the obtaining of satisfactions then was measured in a situation in which the *S* himself has some control over failure or success.

One kind of behavior studied was that of sorting figures taken from the MAPS test. The *S* was presented the figures and asked to sort them into two piles any way he wished. The procedure was repeated a second time asking for a different kind of sort, and a third time. The sorts themselves were characterized as being safe or cautious in

[4] Cautious and noncautious behavior was characterized according to the patterns described by Rotter (1954, pp. 318–324). Patterns 1 and 3 were considered as noncautious and 2, 4, 7, and 8 as cautious patterns. The latter group are characterized by a variety of techniques presumably aimed at avoiding failure to reach explicit goals. Patterns 1 and 3 are characterized by higher expectancies than performance but within "normal" bounds and consequently a higher number of failures to reach one's estimates.

that they dealt with highly objective characteristics of these figures, or less safe in that they dealt with characteristics which were more abstract or had to be read into many of the figures. For example, sorts based on personality characteristics were considered as noncautious as opposed to safer or more cautious sorts such as those into groups of men and women, children and adults, Negroes and whites, etc.

A second kind of behavior studied was the *S*'s response to a series of four TAT pictures. In this case the *S*'s stories were treated as Weisskopf (1950) has treated them with her measure of transcendence. A cautious or safe interpretation was one sticking close to the characteristics of the picture and one in which the theme itself was a common one.

Moss found some differences among his groups in the direction he had hypothesized, that is, that the threatened group, the group that was told that it was at the tenth percentile, showed greater cautiousness than the other two groups. The differences between groups, however, although consistently in the direction he predicted, only approached significance and were not large. However, when Moss divided his *S*s within groups into cautious and noncautious on the basis of their level of aspiration patterns, he found some highly significant differences. The cautious *S*s showed no significant differences among the three conditions. However, the noncautious *S*s showed significant differences between conditions. That is, noncautious *S*s in one condition responded differentially from noncautious *S*s in another condition. These differences were primarily due to greater noncautious behavior in the no-information group. Differences in test behavior between cautious and noncautious *S*s were also highly significant on both tests within the no-information group but not in the other groups.

In spite of the complexity of this study, a few findings seem relatively clear from an analysis of group means as well as significance of differences. *S*s who were cautious on the level of aspiration test, which is a somewhat free situation, were also cautious in the other test conditions. Of course, this does not mean that they were cautious in situations which were not perceived by them as evaluation situations. On the other hand, *S*s who were noncautious on the level of aspiration test appeared to maintain this greater risk taking behavior under test situations where no information about results was given. However, when they were negatively reinforced, they became more cautious and they also did not appear to be different from cautious *S*s under conditions where they were quite successful. Perhaps this is related to the presumed conservatism which follows from success. There was no consistent prediction from the level of aspiration situation to the two "projective tests" which could be made without considering the situation. In at least two of the situations the cautious *S*s were not significantly different from the noncautious *S*s. On the other hand, the psychological situation or the three different situations seemed to have no effect on the cautious *S*s. Only in the interaction of the noncautious *S*s with the situational variables was prediction possible from the level of aspiration test.

A similar type of result to that of Moss was recently reported by Lesser (1959). Lesser found that intercorrelations among various measures of aggression were significantly higher under experimental conditions of low anxiety than under conditions of high anxiety about aggression.

James' results suggest a rather simple relationship between dispositional

and situational variables, but it is clear from the study of Moss and other studies that a simple additive or multiplicative relationship will not always describe the nature of the interaction. An important implication of this principle is that there is a general lack of efficiency in research studies in which only one set of variables, that is only dispositional or situational, are systematically varied, since the conclusions of the two sets of studies cannot be put together in a simple fashion. Unless both kinds of variables are systematically varied *within the same investigation*, both an understanding of the determinants of behavior and the prediction of it may suffer.

A striking example of the importance of studying the effects of dispositional and situational influences simultaneously is provided by Helson, Blake, Mouton, and Olmstead (1956). In applying Adaptation Level theory to a study of attitude change they found important interactional effects when situations were varied in external influence pressure and individuals were distributed on a measure of ascendancy–submissiveness.

It is true that many of the above propositions are obvious. Most psychologists recognize that there is a difference between overt behavior directed toward certain goals and the desires that individuals have to obtain these goals. Similarly, most psychologists know that the psychological situation is a determinant of the occurrence of a given behavior. The thesis here, however, is not merely that this is the case but that all of these variables must be ordered and studied systematically, in order to make predictions.

SUMMARY

The major contention of this paper has been that the prediction of goal directed behavior of human subjects from test procedures has been and will continue to be at an extremely low hit-or-miss level because of inadequate conceptualization of the problem. Findings are frequently not replicatable because of the failure to systematically differentiate behavior, reinforcement value, and expectancy as internal variables and to recognize that these variables are affected by the psychological situation.

The psychological situation of the patient in the clinic is so different from that of the elementary psychology student taking a test as part of an experiment that it is possible that the kinds of predictions which can be made in one situation would hardly hold in the other. The evidence that faking is possible and that different norms obtain when subjects are job applicants, employees, or volunteers does not necessarily mean that a test is no good. Nor is prediction essentially hopeless because it can be demonstrated that two experimenters, whether the same sex or opposite, or slight changes in the wording of instructions, will differentially affect test or experimental results. All of these things indicate only that the psychological situation, perhaps acting primarily through the expectancies they arouse by the cues present, considerably affect behavior. It is necessary that we do not consider such influence as error to be ignored, as difficulty to be avoided or as the problem of some other profession to investigate. Rather it is necesary to study these influences and consider them regularly and systematically in a predictive schema. That is, for some purposes, factors such as social desirability of items, examiner's behavior, and the subject's goals in the test situation should be controlled, and in other cases they should be allowed to vary. In all cases, however, they must be systematically considered.

Implicit in this entire paper is the belief that a satisfactory theory of goal

directed behavior is a primary prerequisite for developing adequate tests. Knowledge of statistics and test construction procedures can be valuable but they cannot supplant an adequate theory of behavior which is applied to the test taking behavior itself.

To arrive at a fully systematic model for relating these general or high order constructs and to coordinate them in turn to lower level sets of content variables, devised for different purposes, will be a long and arduous but rewarding task.

REFERENCES

ALLPORT, G. W. What units shall we employ? In G. Lindzey (Ed.), *The assessment of human motives.* New York: Rinehart, 1958.

ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situation. *J. Pers.*, 1956, **24**, 349–363.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

BORISLOW, B. The Edwards Personal Preference Schedule and fakability. *J. appl. Psychol.*, 1958, **42**, 22–27.

BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. California Press, 1947.

BUTLER, J. M. The use of a psychological model in personality testing. *Educ. psychol. Measmt.*, 1954, **14**, 77–89.

COUTU, W. *Emergent human nature.* New York: Knopf, 1949.

CRANDALL, V. J. Induced frustration and punishment-reward expectancy in thematic apperception stories. *J. consult. Psychol.*, 1951, **15**, 400–404.

CRONBACH, L. J. Assessment of individual differences. In P. R. Farnsworth and Q. McNemar (Eds.), *Annu. Rev. Psychol.*, 1956. Stanford, Calif.: Annual Reviews, 1956. Pp. 173–196.

CRONBACH, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, 12, 671–684.

EDWARDS, A. L. *Manual of the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1953.

FITZGERALD, B. J. Some relationships among projective test, interview and sociometric measures of dependent behavior. *J. abnorm. soc. Psychol.*, 1958, **56**, 199–204.

GIBBY, R. G., MILLER, N. R., & WALKER, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, **17**, 425–428.

GREEN, R. F. Does a selection situation induce testees to bias their answers on interest and temperament tests? *Educ. psychol. Measmt.*, 1951, **11**, 503–515.

GROSS, L. R. Effects of verbal and nonverbal reinforcement in the Rorschach. *J. consult. Psychol.*, 1959, **23**, 66–68.

HELSON, H. Adaptation level as a basis for quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HELSON, H., BLAKE, R. R., MOUTON, JANE S., & OLMSTEAD, J. A. Attitudes as adjustments to stimulus, background, and residual factors. *J. abnorm. soc. Psychol.*, 1956, **52**, 314–322.

HENRY, EDITH M., & ROTTER, J. B. Situational influences on Rorschach responses. *J. consult. Psychol.*, 1956, **20**, 457–462.

HUNT, D. E., & SCHRODER, H. M. Assimilation, failure-avoidance, and anxiety. *J. consult. Psychol.*, 1958, **22**, 39–44.

JAMES, W. H. Internal versus external control of reinforcement as a basic variable in learning theory. Unpublished doctoral dissertation, Ohio State Univer., 1957.

JAMES, W. H., & ROTTER, J. B. Partial and one hundred percent reinforcement under chance and skill conditions. *J. exp. Psychol.*, 1958, **55**, 397–403.

JESSOR, R., & HAMMOND, K. R. Construct validity and the Taylor anxiety scale. *Psychol. Bull.*, 1957, **54**, 161–170.

KAGAN, J. The measurement of over aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, **52**, 390–393.

KANTOR, J. R. *Principles of psychology.* Vols. 1, 2. New York: Knopf, 1924.

LASKO, A. A. The development of expectancies under conditions of patterning and differential reinforcement. Unpublished doctoral dissertation, Ohio State Univer., 1952.

LESSER, G. S. The relationship between overt and fantasy aggression as a function of maternal response to aggression. *J. abnorm. soc. Psychol.*, 1957, **55**, 218–222.

LESSER, G. S. Population differences in construct validity. *J. consult. Psychol.*, 1959, **23**, 60–65.

LEWIN, K. The nature of field theory. In M. H. Marx (Ed.), *Psychological theory.* New York: Macmillan, 1951.

LITTLE, K. B., & SHNEIDMAN, E. S. Congruencies among interpretations of psycho-

logical test and anamnestic data. *Psychol. Monogr.*, 1959, **73**(6, Whole No. 476).

LIVERANT, S. The use of Rotter's social learning theory in developing a personality inventory. *Psychol. Monogr.*, 1958, **72**(2, Whole No. 455).

LOEVINGER, JANE. Theory and techniques of assessment. In P. R. Farnsworth (Ed.), *Annu. Rev. Psychol.*, 1959. Palo Alto, Calif.: Annual Reviews, 1959. Pp. 287–316.

LORD, E. E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **64**(10, Whole No. 316).

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MANDELL, ELIZABETH E. Construct validation of a psychometric measure of expectancy. Unpublished master's thesis, Univer. Colorado, 1959.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MOSS, H. The generality of cautiousness as a defense behavior. Unpublished doctoral dissertation, Ohio State Univer., 1958.

MURRAY, H. A. Toward a classification of interaction. In T. Parsons & E. A. Shils (Eds.), *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1952.

MURSTEIN, B. I. A conceptual model of projective techniques applied to stimulus variations with thematic techniques. *J. consult. Psychol.*, 1959, **23**, 3–14.

MUSSEN, P. H., & NAYLOR, H. K. Relationship between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, **49**, 235–239.

MUSSEN, P. H., & SCUDEL, A. The effect of sexual stimulation under varying conditions on TAT sexual responsiveness. *J. consult. Psychol.*, 1955, **19**, 90.

NAYLOR, H. K. The relationship of dependency behavior to intellectual problem solving. Unpublished doctoral dissertation, Ohio State Univer., 1955.

NEFF, J. Individual differences in resistance to extinction as a function of generalized expectancy. Unpublished doctoral dissertation, Ohio State Univer., 1956.

PEAK, HELEN. Problems of objective observation. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences.* New York: Dryden, 1953.

PHARES, E. J. Expectancy changes in skill and chance situations. *J. abnorm. soc. Psychol.*, 1957, **54**, 339–342.

RORSCHACH, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.

ROTTER, J. B. Level of aspiration as a method of studying personality: II. Development and evaluation of a controlled method. *J. exp. Psychol.*, 1942, **31**, 410–422.

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice Hall, 1954.

ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1955. Lincoln: Univer. Nebraska Press.

SARASON, S. The test situation and the problem of prediction. *J. clin. Psychol.*, 1950, **6**, 387–392.

THOMAS, W. I. (Collected writings) In E. H. Volkart (Ed.), *Social behavior and personality: Contributions of W. I. Thomas to theory and social research.* New York: Social Science Research Council, 1951.

TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.

WEISSKOPF, E. A. A transcendence index as a proposed measure of projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 379–390.

(Received February 2, 1959)

Psychological Review
1960, Vol. 67, No. 5, 317–331

LEARNING THEORY AND THE ACQUISITION OF VALUES[1]

WINFRED F. HILL

Northwestern University

The processes by which a child acquires the values of his culture and his various overlapping subcultures is, as a recent review of the subject points out (Dukes, 1955), still rather obscure. This obscurity is certainly not due to lack of interest in the topic. Psychologists, psychiatrists, anthropologists, sociologists, pediatricians, and educators have all given attention to the question of how children come to share the attitudes, ideals, and ethical standards of those around them. Nor is this interest undeserved, for few topics are of greater practical importance.

Perhaps this very convergence of interest from many directions is partly responsible for the difficulties involved in studying the topic. This area of research has become a battleground for conflicting terminologies, with one term often having a multiplicity of half-distinct meanings, and with what appears to be the same meaning often bearing different labels. Although many terms contribute to this confusion, three are of particular interest here: *identification, introjection,* and *internalization.* All involve some relation between an individual, hereinafter designated the subject (S), and another person or personalized entity, the model or M, such that S's behavior is in some way patterned after M's. However, these terms may refer either to a state of affairs or to the process which brought it about (Lazowick, 1955); the M may be a person, a group, or an idea (Glaser, 1958); and the relation may involve specific responses, broad meanings, or emotional reactions.

Some of the confusion as to the meaning of the term identification may be seen in the following uses. Lazowick (1955) distinguishes three main uses of the term identification in the literature: pseudoidentity, imitation, and personality change. He suggests that the term should be used only with regard to broad meanings, with imitation being the corresponding term for specific acts. Freud (1950) contrasts a boy's identification with his father, which forms the basis of his ego ideal, and his identification with his mother as an abandoned object cathexis. Davis and Havighurst (1947) maintain that a child will identify with his parents only if he loves them, but Anna Freud (1946) emphasizes identification with the aggressor. Lynn (1959) contrasts sex-role identification, which is "reserved to refer to the actual incorporation of the role of a given sex, and to the unconscious reactions characteristic of that role" (p. 127), with sex-role preference and sex-role adoption. He regards figure drawing as a measure of identification and choice of dolls for play as a measure of preference. Sears, Maccoby, and Levin (1957), however, use children's choices of dolls as a measure of identification. Sanford (1955), discouraged by such confusions of meaning, considers the possibility of abandoning the term identification altogether, but decides to retain it to describe a defense mechanism involving

[1] The research for this article was supported by a grant from the Carnegie Corporation to the Department of Psychology and the School of Education at Northwestern University. The author wishes to express his appreciation to Donald Campbell, Robert Winch, and the members of their seminar in social psychology for their many contributions to his thinking on this topic.

extreme adoption of *M*'s behavior by *S*, a mechanism which is not important in normal personality development. Finally, this collection of meanings, diverse as it is, still omits those cases where identification is used as a synonym for loyalty or for empathy.

There is similar confusion concerning the meaning of introjection and internalization. Both carry the implication of values being incorporated into the personality. Hence, particularly with introjection, there is the suggestion of some relation with orality (Freud, 1950). However, Freud in the same discussion also uses introjection synonymously with identification. Parsons (1955), on the other hand, treats identification and internalization as synonyms.

The many discussions of these three terms in the literature seem to indicate that there are several processes involved but no generally accepted conventions for labeling them. A number of writers, including several already cited, have expressed discouragement at this state of affairs, but the usual result of such discouragement seems to be a redefinition of terms, which may clarify the particular exposition but which only serves further to confuse the field as a whole. Whereas Lynn (1959) believes that a term as widespread as identification must have potential usefulness, the present writer believes that clarity would be served by abolishing not only identification but also introjection and internalization from the technical vocabulary of personality development.

What, then, should be substituted? The topics to which the above terms have been applied certainly deserve discussion, and if the redefinition of the old terms is unsatisfactory, the introduction of new terms would be even worse. An answer may be found, however, in a sort of reductionism. Since the processes involved are learning processes, the existing vocabulary of learning is the obvious candidate for the job of describing them. It is quite possible, of course, that the existing vocabulary of learning theory will be inadequate for the complexities of value acquisition. However, if its use is carried as far as possible, the successes of this application should clarify our thinking about personality development, while any gaps which result should point to possible extensions of learning theory.

This approach involves treating human learning in a sociocultural environment in the same terms, at least for a first approximation, as animal learning in the environment of laboratory apparatus. For this purpose, the social rewards and punishments applied to humans may be treated as equivalent to the food pellets and electric shocks used with rats. Similarly, social roles are the equivalents of mazes which must be learned in order to obtain the rewards and avoid the punishments. Human beings of course constitute a far more variable environment than laboratory hardware, and one on which *S* can exercise greater influence. However, since most of the theory in this area is concerned with the adaptation of *S* to a relatively constant human environment (whether it be called culture, social system, or the personalities of the parents), this should not prove a serious stumbling block. There is ample precedent for such an approach in the writings of Dollard and Miller (1950), Mowrer (1950), Whiting and Child (1953), and others.

In addition, this approach treats values as nothing more than inferences from overt behavior. In principle this assumption should cause no difficulty. Few behavioral scientists would regard values (in the empirical, not the transcendental sense) as fundamentally dif-

ferent from such behavioristic constructs as Hull's (1943) habit strength or Tolman's (1949) equivalence beliefs. In practice, however, some theorists might take issue with this view on at least two bases.

For one thing, the measurement of values (including attitudes, ideals, and ethical standards) is commonly by verbal methods (see the review by Dukes, 1955). This leads to the suspicion that any measurement of values by nonverbal means must be inadequate, that only verbal measures can get at the significance of an act for an individual. However, verbal responses are part of the total behavior of the human organism and may be studied like other responses. The processes of unconscious motivation, semiconscious hypocrisy, and deliberate concealment all indicate that it would be unwise to treat verbal and nonverbal measures of values as equivalent. Rather than treat either verbal or nonverbal behavior as the true indicator of values and the other as a side issue, it seems more useful to study both and to ascertain empirically to what extent they lead to the same generalizations about a given *S*.

Another possible objection involves the distinction between specific acts and broad meanings, as in Lazowick's (1955) contrasting definitions of imitation and identification, noted above. This distinction between specific acts and broad meanings may refer either to the presence of mediating responses (see Osgood, 1953) or to the generality of the stimuli and responses involved. Neither of these distinctions, however, is dichotomous. Hull (1952) has indicated how a mediator, the fractional antedating goal response, may function in animal behavior, and Russell and Storms (1955) have demonstrated mediational processes in rote learning. Thus the mediation mechanism is by no means restricted to the "higher mental processes" of humans. As for the generality of the behavior and of the stimuli which guide it, this presumably represents a continuum from the most specific to the most inclusive categories. If, for example, washing the hands before meals is an example of imitation and cleanliness an example of identification, where would wearing clean clothes be classified? So, although the distinction between specific acts and broad meanings is a legitimate one, there are no sharp breaks on the continuum and there is no reason to assume that basically different laws are involved.

In view of the above considerations, an attempt to study the acquisition of values as a branch of learning theory appears justified.

KINDS OF REINFORCEMENT

The concept of reinforcement is basic to learning theory. While theorists are by no means unanimously agreed on the value of reinforcement terminology, there is little question that an empirical law of effect holds, that the consequences of an act influence its subsequent occurrence. A classification of kinds of reinforcers will be used here as the basis for analyzing the learning of values.

Primary Reinforcement

For the present purpose, three kinds of reinforcement may be distinguished: *"primary," secondary,* and *vicarious*. Placing "primary" in quotes indicates that it refers to the effects not only of innate physiological reinforcers but also of those social reinforcers which play a primary role in human motivation. Presumably the positively reinforcing effects of attention and praise and the negatively reinforcing effects of criticism, ridicule, and rejection are at least partly learned, but the nature of the

learning process is obscure, and at the present level of analysis it seems preferable to treat praise for a human as comparable to food for a rat. The distinction between "primary" and secondary reinforcement is thus one of convenience between that which we take as given and that for which we can find a specific learned basis. Though arbitrary, this distinction is perhaps no more so than the decision as to whether food in the mouth should be considered a primary or a secondary reinforcer.

One particular kind of learning by "primary" reinforcement is the acquisition of a generalized tendency to imitate others. Miller and Dollard (1941) have indicated how a generalized tendency to imitate the behavior of others may be learned in the same way as any other class of responses. Although their demonstrations of imitation involved *S*'s patterning his behavior after a leader who was present, Church (1957) has shown that rats can also learn to respond appropriately to the same cues to which the leader rat is responding. In spite of some negative animal evidence (Solomon & Coles, 1954), there is little doubt that humans can learn to pattern their behavior after that of other people, not only when the *M* is present, but also in *M*'s absence by utilizing the appropriate environmental cues. As the child is repeatedly rewarded for imitative behavior in a variety of otherwise different situations, and as his capacity for abstraction increases, it seems plausible that a generalized imitative tendency would develop. It would be desirable, however, if the widespread anecdotal support for this deduction could be bolstered by experimental data.

The same process presumably applies to verbal instructions. The child is typically reinforced (though with some striking exceptions) for doing what others tell him to do. Hence (common parental impressions to the contrary notwithstanding), a generalized tendency toward conformity to verbal instructions may be expected to develop. With increasing intellectual development, this tendency should come to include conformity to fictional examples or to abstract ethical exhortations.

Secondary Reinforcement

Although no basic distinction is made here between primary and secondary reinforcement, there is one case frequently discussed in the literature where the acquisition of reinforcing properties by certain stimuli may be analyzed in detail. These stimuli are those which are connected with care of the child by adults, i.e., the nonessential aspects of nurturance. These include patterns of speech, facial expressions, gestures, and the like. Since these occur with those nurturant behaviors which are primary reinforcers, such as feeding and cuddling, they may become secondary reinforcers. By stimulus generalization, these behaviors should also be rewarding to the child (although less so) when produced by himself. As the child grows older and the parents expect him to take greater care of his own needs, he is more and more forced to provide not only his own primary reinforcers, but his own secondary reinforcers as well. Hence he may be expected to show some of the same mannerisms as his parents showed when caring for him.

This kind of learning appears to be one of those processes which Freud (1950) includes under the heading of identification or introjection, that in which abandoned object cathexes become incorporated into the ego. However, in the view presented here, the coincidence of abandonment and incorporation into the ego refers only to performance, not to learning. The

secondary reinforcing value of the parental mannerisms is built up during the period of nurturance, but becomes evident in the child's behavior as nurturance begins to be withdrawn. This learning process has been discussed by Mowrer (1950, Ch. 24) in connection with the learning of language, and by Lindemann (1944) as a reaction to the death of a loved one. Although this process appears better adapted to the learning of rather trivial mannerisms, it is capable at least in principle of being adapted to more general and significant values as well.

Vicarious Reinforcement

Vicarious reinforcement does not have the same dignified status in learning theory as do primary and secondary reinforcement, but some such process appears necessary in order to explain some important human learning. Vicarious reinforcement involves the generalization of reinforcing effects from others to oneself, hence learning from the reinforcers which others receive. A given act is reinforced for S as a result of the act being performed by M, followed by reinforcement to M. For example, if S observes M trying to solve a problem by certain techniques and succeeding, S is more likely to use the same techniques when faced by a similar problem than if M had failed to solve the problem. Although most of the evidence for such learning is anecdotal, Lewis and Duncan (1958) have provided some evidence of it in a human gambling situation, and Darby and Riopelle (1959) have demonstrated it in discrimination learning by monkeys.

Although vicarious reinforcement involves selective imitation, it differs from the selectivity of imitation, described by Miller and Dollard (1941), in which S imitates some Ms and not others because of differential reinforcement received by S for imitating the two Ms. Vicarious reinforcement does not involve any reinforcers delivered directly to S; the discrimination of Ms to be imitated or not is made entirely on the basis of S's observation of M's experience. This distinction is emphasized by Campbell (in press). In Hullian terms, vicarious reinforcement involves the acquisition of K by observation.

This type of learning need not be restricted to the effect of particular reinforcers administered to M under specific conditions. Stimulus generalization should occur not only from M's behavior to S's but also from one act of M's to another. As a result, if M is frequently reinforced, S should find it rewarding to resemble M in general, including imitation of some of M's behaviors which S has never seen rewarded. Thus a beginning salesman (S) might treat a customer with extreme politeness because he had observed another salesman (M) making a large sale while using such behavior, and he might also smoke a cigar because he had observed his highly successful salesmanager (M) doing so. In the former case M's behavior (politeness) and M's reinforcement (a large sale) were paired, whereas in the latter case the reinforcement (business success) was a perennial experience of M, but not paired with the particular behavior (smoking cigars) in question. Both, however, are examples of vicarious reinforcement.

Vicarious reinforcement corresponds to identification as defined by Masserman (1946) and to that aspect of identification referred to by Kagan (1958) as "the motivation to command or experience desired goal states of a model" (p. 298). Freud's (1946) identification with the aggressor also fits under this heading if successful aggression is assumed to be reinforcing

to the aggressor. Sanford's (1955) concept of identification also involves the adoption as *M* of someone perceived by *S* as successful.

Conflicting Sources of Reinforcement

Traditionally the terms identification, introjection, and internalization might be applied to any or all of the learning processes described above or to their end product, similarity between *S* and some *M*. Since for the most part these processes involve learning by imitation, require some kind of reinforcement, and result in similarity between *S* and *M*, it may be asked why detailed analysis of the rather subtle differences among them is called for.

The answer is apparent when the possibility of conflict is considered. Conditions for a given *S* may be such that one of these processes tends to produce one kind of behavior while another tends to produce quite different or even opposite behavior. Such a conflictful situation might be expected in a child reared by a nurturant mother, whose mannerisms would become secondarily reinforcing, and a domineering father, who would be perceived as successful in mastering the environment. Freud (1950) recognized the frequent occurrence of just such a conflict, but did not consider it necessary to use different words for the two learning processes. Another common conflict is between the tendency to imitate *M*s whom *S* is directly reinforced for imitating (e.g., well behaved children) and the tendency to imitate *M*s whom *S* perceives as successful (e.g., tough kids). Such conflict is inevitable to a certain extent in children, since they are not permitted to imitate their (presumably more successful) elders in all respects, but it is particularly prominent in members of low-status social categories, (e.g., Negroes) who are often conspicuously not reinforced for imitating high-status *M*s. In the broadest sense, any situation where there is discrepancy among what *S* is told to do, what he is rewarded for doing, and what he sees others doing is a potential conflict situation, and one in which the use of any single inclusive term such as "identification" obscures the relevant variables.

The occurrence of conflict among the various reinforcement processes makes possible a finer analysis of the acquisition of values than could be made otherwise. If there is perfect agreement among what *S* is told to do, what those who nurture him do, what those around him conspicuously master the environment by doing, and what he himself is directly rewarded for doing, there is little basis for judging how much each of these factors contributed to *S*'s adoption of the values of those around him. By observing situations in which they conflict, greater knowledge of the efficacy of each kind of reinforcement may be obtained. Research in conflict situations might answer such theoretically and practically important questions as: "Does dominance or nurturance on *M*'s part do more to make *M* effective in modifying *S*'s values?"; "Do words and examples completely lose their efficacy if the appropriate behaviors, when elicited, are not reinforced?"; and "To what extent is behavior influenced by *M*s presented verbally (e.g., in literature)?"

There is of course no special merit to the classificatory scheme presented here. Except for the concept of vicarious reinforcement, the writer has avoided attaching distinctive labels to the learning processes described. The purpose of this discussion was to show how the terminology of learning theory can be applied to processes of value acquisition which have been described by personality theorists. This

not only serves as a step toward the integration of these two areas of study, but also suggests the probable usefulness of employing such independent variables as number, percentage, magnitude and delay of reinforcement, distribution of practice, and discriminability of stimuli in the study of value acquisition. As both learning theory and personality theory develop further, it is to be expected that any schema developed now will be at least partially replaced by newer concepts. Rather than developing in further detail the ideas suggested above, the remainder of this discussion will therefore concentrate on the application of this kind of thinking to a narrower area, the development of conscience.

CONSCIENCE

Negative values, or conscience, have received much more attention than positive values. Educators seeking to improve children's characters, psychoanalysts concerned with the tyranny of the superego, anthropologists trying to distinguish between shame and guilt cultures, and experimental psychologists noting the persistence of avoidance responses have shared this emphasis on values of the "Thou shalt not" variety. Because of this widespread interest in conscience, it is a particularly appropriate topic with which to illustrate the possibilities of the learning theory approach to the study of values. Sears, Maccoby, and Levin (1957), in their challenging book *Patterns of Child Rearing*, devote a chapter to the development of conscience in preschool children. Their treatment of the topic will serve as a starting point for the present analysis.

Criteria for Conscience

Sears, Maccoby, and Levin give three criteria for recognizing the operation of conscience in young children: *resistance to temptation, self-instructions* to obey the rules, and *evidence of guilt* when transgression occurs. These three criteria are treated jointly as defining conscience, and no attempt is made to analyze their separate developments. Although the authors mention that the aspects of conscience do not necessarily all appear at once, they regard conscience as representing an internalization of control which is fundamentally different from external control, whether by force, fear of punishment, or hope of material reward. This treatment of conscience as essentially a single variable seems premature in our present state of knowledge; certainly the learning theory approach to personality advocated here would involve separate analyses of these diverse response patterns.

The first criterion, *resistance to temptation*, may be viewed simply as avoidance learning. Although this kind of learning is still a focus of theoretical controversy, much experimental data are available concerning it (Solomon & Brush, 1956). Sidman's (1953) studies of avoidance behavior without a warning signal and Dinsmoor's (1954) analysis of punishment show how feedback from an individual's own acts can become a cue for avoidance, and how persistent such avoidance may be. Although children can presumably learn to respond to more abstract characteristics of cues than can animals, there is no reason to regard a child's learning to avoid certain behaviors as fundamentally different from a rat's learning to do so. The fact that the child avoids the forbidden acts even in the absence of the parents is presumably due to the parents' having in the past discovered and punished (in the

broadest sense of that word) transgressions committed in their absence.

This relating of conscience to avoidance learning suggests that independent variables known to be effective in animal avoidance learning would be among the most appropriate ones for study in connection with the development of conscience in children. Within certain limits, the greater the intensity of the punishments (Brush, 1957; Miller, 1951) and the shorter the delay between transgression and punishment, (Mowrer & Ullman, 1945; Solomon & Brush, 1956) the greater should be the resulting inhibition. Though the data are somewhat ambiguous, greater certainty of punishment might be expected to produce inhibition which would be more complete in the short run but also less persistent once punishment was permanently withdrawn (Grant, Schipper, & Ross, 1952; Jenkins & Stanley, 1950; Reynolds, 1958). This prediction suggests that even this one criterion of conscience may not be unitary, that different laws may apply depending on whether one asks how completely the child obeys the prohibitions or how long he continues to obey them after leaving the parental home. If partial reinforcement should turn out to be a crucial variable in the human situation, these two criteria might even be inversely related. The prediction also suggests that the question, "Is inconsistent discipline bad?" is far too simple; one must at least ask, "Bad for what?"

It must also be kept in mind that punishment is not restricted to physical chastisement or even to noxious stimuli in general, including scolding and ridicule. Withdrawal of positive reinforcers may be very effective as a punishment, a fact which complicates the analysis. As this is a much discussed topic in personality theory, it will be considered below.

Sears, Maccoby, and Levin's second criterion of conscience, *self-instruction*, obviously makes the human case different from the animal case, but it does not introduce any new motivational principle. One of the advantages of membership in the human species is the possibility of using verbal symbolization in dealing with one's problems. It is natural that a person learning an avoidance, like a person learning any other difficult response pattern, should give himself verbal instructions, especially since verbal coaching by others is so important in the learning of social prohibitions. Moreover, such self-instruction is an imitative act which might be learned according to any of the reinforcement paradigms discussed above. Presumably the learning of prohibitions proceeds differently in verbal and nonverbal organisms, but observations of the relation between moral statements and moral behavior (Hurlock, 1956, pp. 406, 411–412) argue against the assumption that there is a high correlation between verbal and other criteria of conscience, except as both are influenced by the values represented in the social environment.

The third criterion of conscience, *guilt* at violations of the prohibitions, is itself complex, with many verbal, autonomic, and gross behavioral aspects. However, the striking paradox about guilt, which has seemed to some students to set it apart from the ordinary laws of learning, is that it often involves the seeking of punishment. The person who has transgressed, rather than trying to avoid punishment, or even waiting passively for it to come, actively seeks out the authorities, confesses, and receives his punishment with apparent relief. He may

also, or instead, go to great lengths to make restitution. Were it not for these phenomena of punishment-seeking and self-sacrificing restitution, it would be easy to dismiss guilt as merely the kind of fear associated with anticipation of certain sorts of punishment. As it is, the existence of guilt serves as an argument for regarding conscience as something more than the sum of all those avoidances which have moral significance in one's culture.

However, the attempt to distinguish between guilt-controlled and other behaviors has not been very successful. Though the distinction between guilt cultures and shame cultures has had a considerable vogue in anthropology (e.g., Benedict, 1946; Havighurst & Neugarten, 1955; Mead, 1950), the inadequacies of the distinction have been pointed out by Ausubel (1955) and by Singer (1953). Moreover, the relation between conformity to a standard and guilt when the standard has been violated is open to question. Shaw (1948) suggests that confession may even be so satisfying to some people that it constitutes a reinforcement for sinning. So, although the phenomena of guilt may raise difficulties for learning theory, these difficulties probably cannot be solved by using guilt to define a distinctive kind of learning.

The above considerations should suffice to indicate that conscience cannot be assumed a priori to be unitary. The extent to which short-run conformity, long-run conformity, self-instructions to conform, certain kinds of distress at having failed to conform, and voluntary confession of nonconformity are intercorrelated is a matter to be empirically determined. Moreover, even if high positive intercorrelations are found, it is possible that they may reflect correlations in the environment rather than any fundamental unity of process. If environmental pressures toward conformity vary markedly, artificially high correlations among the criteria of conscience are to be expected. However, even when this artifact is removed, an analysis of separate learning processes for different behaviors may still lead to the prediction of high correlations among the behaviors. Such an analysis is presented below.

Learning of Conscience

Sears, Maccoby, and Levin found that the development of conscience, as defined jointly by their three criteria, was greater in those children whose parents used love-oriented forms of discipline (praise, isolation, and withdrawal of love) than in those whose parents used "materialistic" forms of discipline (material rewards, deprivation of privileges, and physical punishment). A similar finding, though not highly reliable statistically, is reported by Whiting and Child (1953, Ch. 11) in a cross-cultural study of guilt as measured by attitudes toward illness. This is consistent with the widely held view that the acquisition of parental values occurs most fully in an atmosphere of love (e.g., Ausubel, 1955; Davis & Havighurst, 1947). It is possible, however, that this finding may be due, not to love-oriented discipline as such, but to other characteristics of discipline which are correlated with it. The effect of this kind of discipline may be to accentuate the learning of several different responses, all of which contribute to the overall diagnosis of high conscience.

The various kinds of punishments commonly applied to children probably differ markedly in the temporal relations and the reinforcement contingencies involved. Physical punish-

ment is likely to occur all at once and be over quickly, while punishment by deprivation of objects or privileges is likely to be either for a fixed period of time or for as long as the disciplinarian finds convenient. Discipline by withdrawal of love, on the other hand, probably much more often lasts until the child makes some symbolic renunciation of his wrongdoing, as by apologizing, making restitution, or promising not to do it again. The child is deprived of his parents' love (or, the parents would claim, of the outward manifestations of it!) for as much or as little time as is necessary to get him to make such a symbolic renunciation. When he has made it, he is restored to his parents' favor. If the normal relation between the parents and child is one of warmth, such discipline strongly motivates the child to make the renunciation quickly. On repeated occasions of transgression, punishment by withdrawal of love, and symbolic renunciation, the child may be expected not only to learn the renunciation response as an escape from parental disfavor but eventually to use it as an avoidance rather than merely an escape response. Thus if the wrongdoing is not immediately discovered, the child may anticipate his parents' impending disfavor by confessing in advance and making the symbolic renunciation.

The result of this hypothesized sequence of events is that the child makes a verbal response which is in effect an instruction to himself not to repeat his wrongdoing. The next time temptation comes, he is more likely to make this verbal response before transgressing. Although this does not guarantee that he will not transgress, it is likely to reduce the probability. If he succumbs to temptation, he is more likely to confess before being caught and thereby avoid the temporary loss of his parents' love. Thus if the above reasoning is correct, all three criteria of conscience should be present to a greater degree in the child who has been disciplined in this fashion than in other children. According to the present hypothesis, however, this will be due to the fact that punishment continues until the child makes a symbolic renunciation, rather than to the fact that the punishment involves withdrawal of love. If physical chastisement or loss of privileges are used in the same way, the same outcome is predicted.

A possible weakness of this hypothesis is that children might learn a discrimination between the symbolic and the actual avoidances, so that they would develop a pattern of violating parental standards, immediately confessing and apologizing, and then transgressing again at the next hint of temptation. If forgiveness is offered freely and uncritically enough, such a pattern presumably does develop. In this case the correlation among the criteria of conscience would be expected to drop, actual avoidance of wrongdoing no longer being associated with the other criteria. (For this reason, Sears, Maccoby, and Levin might have found smaller relations if they had studied older children.) However, if the parents' discrimination keeps up with the child's so that the child cannot count on removing all the parents' disfavor with a perfunctory apology, the efficacy of this kind of discipline should be at least partially maintained.

If this explanation of greater conscience in children disciplined by withdrawal of love is correct, why was greater conscience also found with the other kinds of love-oriented control? Since these were all found to be intercorrelated, and since their relations to the degree of conscience were uni-

formly low, interpretations either of separate techniques or of love orientation as a general trait are necessarily somewhat dubious. As an example of the difficulties involved, it may be noted that reasoning with the child is counted as a love-oriented technique solely on the grounds of its correlation with the other such techniques. Nevertheless, it shows a higher relation to conscience than do two of the three clearly love-oriented techniques. In view of such complexities, it seems legitimate to suggest that the crucial factor in those techniques associated with conscience may not be love orientation as such, but something else correlated with it.

To test this hypothesis, it would be necessary to have further detailed information of the sort that Sears, Maccoby, and Levin used, so that disciplinary methods could be classified according to the time relations discussed above. It is predicted that the parents' tendency to make termination of punishment contingent on symbolic renunciation would be correlated with love-oriented discipline. However, if each were varied with the other held constant, conscience should be more closely related to response contingency than to love orientation.

Along with this overall analysis of conscience, more detailed analyses could be made of the various components of conscience. According to the present view, intercorrelations among these criteria would be moderate for the entire sample and low when method of discipline was held constant.

The learning sequence discussed above is only one of several possible explanations of the Sears, Maccoby, and Levin finding. By suggesting that the crucial causal factor is not the distinction between materialistic orientation and love orientation, but another distinction correlated with it, the present hypothesis gains an advantage in objectivity and in practical applicability. Whether it also has the advantage of correctness must be empirically determined. The chief purpose of the present example is to point to the availability of such reductionist hypotheses in the study of values and to argue that they deserve priority in the schedule of scientific investigation.

Permanence of Conscience

It would be particularly desirable to have a follow-up study to compare evidences of conscience in kindergarten with those of the same people later in life, when they were no longer primarily under the direct influence of their parents. Such a follow-up would help to clarify the relation between short-run and long-run conformity discussed above. Is the child who thoroughly obeys all his parents' prohibitions also the one who sticks to these standards when his parents are no longer around and his new associates have different standards? Anecdotal evidence can be cited on both sides, though the bulk is probably in the affirmative. To the extent that current and later conformity are independent, what variables influence one more than the other?

Predictions from learning theory on this topic are by no means unambiguous. Nevertheless, two lines of reasoning may be suggested concerning the type of discipline likely to result in the most persistent avoidances. (Persistence here refers, not to absolute level of avoidance, but to relative lack of decrement in the strength of avoidance with time.)

The first line of reasoning is from the differences in the slopes of *generalization gradients* for different

kinds of learning (Dollard & Miller, 1950). In most cases the contrast in slope is between approach (or excitatory) and avoidance (or inhibitory) tendencies. It appears, however, that the basic distinction is between response tendencies activated by innate and by learned (generally fear) drives (Miller & Murray, 1952). When stimulus conditions change, the resultant removal of cues for fear produces a greater weakening of response tendencies based on fear than of response tendencies based on other drives. Hence, the generalization gradient of responses and inhibitions based on fear is steeper than that of other responses and inhibitions. This implies that discipline based on fear should lose its efficacy more quickly than discipline based on rewards as distance from the disciplinarian or any other change in conditions increases. Since this difference in slope is found on continua both of distance (Miller, 1944) and of similarity (Miller & Kraeling, 1952; Murray & Miller, 1952), it seems reasonable to predict that it also applies to that complex continuum along which an individual makes the transition away from parental apron strings. It would follow from this analysis that of two inhibitions learned in childhood, equal in age and original strength, one learned from the threat of losing rewards would be more effective later in life than one learned from the fear-provoking threat of punishment.

In this analysis, the advantages of discipline by manipulation (including withdrawal) of reward would apply to any kind of reward, material or social, not merely to parental love. However, the desire to continue receiving love from the parents may persist after the child has outgrown the need for other parental rewards, such as gifts and privileges. Discipline by withdrawal of love, in an atmosphere of warmth, might therefore be even more effective than other forms of discipline by denial of reward in producing persistent avoidances.

The other line of reasoning, involving the *partial reinforcement effect*, argues for the persistence of conscience learned by the process outlined above, in which a symbolic renunciation of wrongdoing terminates punishment. Although the greater resistance to extinction of responses which have received less than 100% reinforcement has been demonstrated primarily with positive reinforcement, it applies to negative as well (Humphreys, 1939, 1940; Grant, Schipper, & Ross, 1952). Partial reinforcement is of course present with all kinds of discipline, since punishment depends on the parents' moods and on the social situation, as well as on the child's being caught. However, the above analysis of the kind of punishment which terminates when the child makes a symbolic renunciation of wrongdoing suggests that such discipline may involve an additional source of partial reinforcement. As was indicated above, the child may learn that he can avoid punishment by confessing and apologizing. When this happens, the avoidance starts to extinguish. However, the discerning parent then learns not to accept the apology, and the child is punished anyway. The child must then make a more vigorous and convincing symbolic renunciation than before in order to terminate the punishment. In addition, the discrimination he has made between the symbolic renunciation and the actual avoidance is broken down; punishment can only be prevented by actual avoidance of wrongdoing. If, however, after a period of obedience he once more transgresses and then confesses, he is likely again not to be

punished. This starts the cycle of extinction and reconditioning of the avoidance response going again, thus continuing to provide a reinforcement schedule in which only part of the child's transgressions are punished.

To predict that such partial reinforcement will retard extinction is admittedly problematic, both because of the complexity of the avoidance paradigm and because the unpunished transgressions are assumed to occur in blocks rather than randomly. Nevertheless, the hypothesis deserves consideration, not only as a prediction from learning principles to personality, but also as a case where the needs of personality theory might guide research in learning.

Although these two lines of reasoning agree in predicting maximally persistent conformity to parental prohibitions by children reared in an atmosphere of parental warmth and disciplined by withdrawal of love, they differ in their other predictions. To test these various hypotheses separately would require both short-run and long-run analyses of the effects of a variety of parental discipline patterns. The following hypotheses might be tested: (*a*) that discipline by deprivations (whether of things, privileges, or love) has more persistent effects than discipline by noxious stimulation (whether physical or social); (*b*) that where the child is taught to confess and apologize for his transgressions, avoidance behavior will go through cycles of extinction and reconditioning; and (*c*) that punishing only part of a child's transgressions results in more persistent obedience than does punishing all of them.

SUMMARY AND CONCLUSIONS

It is suggested that the terms identification, introjection, and internalization be replaced by detailed analyses in learning-theory terms of the acquisition of values. A reinforcement framework for such analyses is outlined, and examples are presented dealing with the concept of conscience and the factors influencing its development. It is argued that this would simplify terminology, encourage more precise study, and further the integration of learning and personality theories.

This analysis, like all attempts to integrate the harder-headed and the softer-hearted portions of behavioral science, is open to attack from both sides. On the one hand it may be objected that the present treatment is too cavalier with the interpersonal and intrapsychic complexities of personality development, that the internalization of values and the identification of one person with another cannot be treated as though they were nothing but the simple learning of a rat in a maze. The answer to this objection is that no "nothing but-ism" is intended; it is an empirical matter both to determine how far the principles of learning (not necessarily simple) can go in explaining personality development and to decide how much the additional principles suggested by some writers actually contribute to our understanding of the phenomena in question. The attempt to catch too much complexity at a single stroke may retard rather than advance our understanding.

On the other hand it may be objected that the interpretations given here are untestable, that the variables involve such diverse and subtle behaviors over such long periods of time as to defy adequate measurement. Admittedly the questionnaire, interview, and brief-observation techniques used in this area leave much to be desired. However, as long as applied

behavioral scientists are called upon to deal with questions of personality development, poor data to guide their decisions are better than none. Study of learning of values by humans, guided by the principles of learning based on both animal and human studies, has the potential to make vital contributions to many theoretical and applied areas of knowledge. It is hoped that the present discussion may contribute something to that goal.

REFERENCES

AUSUBEL, D. P. Relationships between shame and guilt in the socializing process. *Psychol. Rev.*, 1955, **62**, 378–390.

BENEDICT, RUTH. *The chrysanthemum and the sword.* Boston: Houghton Mifflin, 1946.

BRUSH, F. R. The effects of shock intensity on the acquisition and extinction of an avoidance response in dogs. *J. comp. physiol. Psychol.*, 1957, **50**, 547–552.

CAMPBELL, D. T. Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 6. *Investigations of man as socius: Their place in psychology and the social sciences.* New York: McGraw-Hill, in press.

CHURCH, R. M. Transmission of learned behavior between rats. *J. abnorm. soc. Psychol.*, 1957, **54**, 163–165.

DARBY, C. L., & RIOPELLE, A. J. Observational learning in the rhesus monkey. *J. comp. physiol. Psychol.*, 1959, **52**, 94–98.

DAVIS, W. A., & HAVIGHURST, R. J. *Father of the man.* Boston: Houghton Mifflin, 1947.

DINSMOOR, J. A. Punishment: I. The avoidance hypothesis. *Psychol. Rev.*, 1954, **61**, 34–46.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

DUKES, W. F. The psychological study of values. *Psychol. Bull.*, 1955, **52**, 24–50.

FREUD, ANNA. *The ego and the mechanisms of defense.* New York: Int. Univer. Press, 1946.

FREUD, S. *The ego and the id.* London: Hogarth, 1950.

GLASER, D. Dynamics of ethnic identification. *Amer. sociol. Rev.*, 1958, **23**, 31–40.

GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. Effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

HAVIGHURST, R. J., & NEUGARTEN, BERNICE L. *American Indian and white children.* Chicago: Univer. Chicago Press, 1955.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

HUMPHREYS, L. G. Psychogalvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–75.

HURLOCK, ELIZABETH B. *Child development.* (3rd ed.) New York: McGraw-Hill, 1956.

JENKINS, W. O., & STANLEY, J. C. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

KAGAN, J. The concept of identification. *Pyschol. Rev.*, 1958, **65**, 296–305.

LAZOWICK, L. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51**, 175–183.

LEWIS, D. J., & DUNCAN, C. P. Vicarious experience and partial reinforcement. *J. abnorm. soc. Psychol.*, 1958, **57**, 321–326.

LINDEMANN, E. Symptomatology and management of acute grief. *Amer. J. Psychiat.*, 1944, **101**, 141–148.

LYNN, D. B. A note on sex differences in the development of masculine and feminine identification. *Psychol. Rev.*, 1959, **66**, 126–135.

MASSERMAN, J. H. *Principles of dynamic psychiatry.* Philadelphia: Saunders, 1946.

MEAD, MARGARET. Some anthropological considerations concerning guilt. In M. L. Reymert (Ed.), *Feelings and emotions: The Mooseheart symposium.* New York: McGraw-Hill, 1950.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 431–465.

MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

MILLER, N. E., & KRAELING, DORIS. Displacement: Greater generalization of approach than avoidance in a generalized approach-avoidance conflict. *J. exp. Psychol.*, 1952, **43**, 217–221.

MILLER, N. E., & MURRAY, E. J. Displacement and conflict: Learnable drive as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.*, 1952, **43**, 227–231.

MOWRER, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.

MOWRER, O. H., & ULLMAN, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, **52**, 61–90.

MURRAY, E. J., & MILLER, N. E. Displacement: Steeper gradient of generalization of avoidance than of approach with age of habit controlled. *J. exp. Psychol.*, 1952, **43**, 222–226.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

PARSONS, T. Family structure and the socialization of the child. In T. Parsons & R. F. Bales (Eds.), *Family, socialization and interaction process.* Glencoe: Free Press, 1955.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *J. exp. Psychol.*, 1958, **55**, 335–341.

RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

SANFORD, N. The dynamics of identification. *Psychol. Rev.*, 1955, **62**, 106–118.

SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

SHAW, G. B. Preface to "Androcles and the Lion." In *Nine Plays.* New York: Dodd, Mead, 1948.

SIDMAN, M. Two temporal parameters of the maintenance of avoidance behavior by the white rat. *J. comp. physiol. Psychol.*, 1953, **46**, 253–261.

SINGER, M. B. Shame cultures and guilt cultures. In G. Piers & M. B. Singer (Eds.) *Shame and guilt.* Springfield, Ill.: Charles C Thomas, 1953.

SOLOMON, R. L., & BRUSH, ELEANOR S. Experimentally derived conceptions of anxiety and aversion. In *Nebraska symposium on motivation IV.* Lincoln: Univer. Nebraska Press, 1956.

SOLOMON, R. L., & COLES, M. R. A case of failure of generalization of imitation across drives and across situations. *J. abnorm. soc. Psychol.*, 1954, **49**, 7–13.

TOLMAN, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, **56**, 144–155.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality.* New Haven: Yale Univer. Press, 1953.

(Received August 3, 1959)

Psychological Review
1960, Vol. 67, No. 5, 332–342

WORD VALUES, WORD FREQUENCY, AND VISUAL DURATION THRESHOLDS

RONALD C. JOHNSON, CALVIN W. THOMSON, AND GERALD FRINCKE

San Jose State College

A number of studies have shown a relation to exist between word frequency, word value, and visual duration thresholds. Certain issues have developed concerning the interpretation of the observed relations. Two types of interpretation of results can be distinguished:

1. Those interpretations that claim word frequency to be the major determinant of visual duration threshold. This point of view contends that the responses made to the tachistoscopic presentation of words are learned in the same manner that other responses are learned. Differences in the visual duration thresholds of words can then be accounted for in terms of word frequency so that the introduction of such tenuous and "unanchored" variables as perceptual selectivity and perceptual defense is a violation of the law of parsimony. Thus, the problems that exist with regard to the differential visual duration thresholds of words are problems in learning, not problems in perception.

2. Those interpretations that attempt to show that differences in the visual duration thresholds of words are due, all or in part, to differences in the affective qualities (values, goodness–badness, affective tone, emotional valence) attached to words. These affective qualities affect visual duration thresholds directly, through perceptual selectivity and/or perceptual defense. Visual duration thresholds of words are determined by variables that can be called perceptual variables instead of or along with those variables that usually bear upon learning.

We wish to center our introductory discussion around those few studies which, for us, best exemplify the two positions described above. Solomon and Howes (1951) take the position that differences between words can be accounted for on the basis of word frequency. The perceptual process does not differ in any fundamental way from the learning process. They say:

> Emotional factors undoubtedly operate to an important extent in the building of word frequencies in a given life history. In this way they would be related to word frequency and, indirectly, to the duration thresholds. . . . But to date we can find no evidence to suggest that emotional factors operate in the tachistoscopic situation independently of their effect on word frequency (p. 267).

According to Solomon and Howes (1951, p. 258) visual duration threshold can be accurately predicted from a knowledge of population-wide word frequency as given in the Thorndike-Lorge tables (1944). There is no need to bring in concepts such as perceptual selectivity or perceptual defense in order to explain differences in visual duration thresholds of words. To Solomon and Howes emotional factors operate only in producing idiosyncratic variation in word frequency. These differences in frequency account for differences in the visual duration thresholds of words such as those found between the visual duration threshold of a value-oriented word for subjects (*S*s) who score high as opposed to low on specific scales of the Allport-Vernon Study of Values. They do not have any other influence on visual duration threshold.

The second position concerning the relation between word frequency, word value and visual duration threshold is clearly expressed by Postman and Schneider (1951). Faced with the problem of salvaging value as a variable influencing visual duration threshold, they say that

> It may . . . be more profitable theoretically to regard both frequency of word usage and duration thresholds as dependent variables, both manifestations of more fundamental psychological properties attributed to the organism, such as "habits," hypotheses, or even, perhaps, "personal values" (p. 277).

While not denying that word frequency can account for a considerable proportion of the variance in visual duration thresholds of words, they claim that values (called directive factors in this paper) directly influence the visual duration thresholds of infrequent words.

Postman, in a later article (1953), suggests that the frequency of usage of a specific word is determined by its reinforcing qualities. He suggests that frequent as opposed to infrequent words differ systematically in affective tone. He presents evidence to show that the sheer number of pleasant words is much greater than that of unpleasant words. Postman says, "Social control over environmental stimuli will, then, tend to create a correlation between frequency and positive value" (p. 68). Whether called values, directive states, or affective qualities, certain emotive or affective aspects of words influence their general (e.g., Thorndike-Lorge) frequency. Affective quality determines frequency, not merely in an idiosyncratic way as Solomon and Howes (1951) suggest, but also in a very general sense for all words used by all individuals.

It has been hypothesized from this second position that values can influence visual duration threshold in two ways. The affective qualities of words influence visual duration threshold *directly* as in the case of an *S*'s Allport-Vernon score influencing his thresholds for infrequent value oriented words (Postman & Schneider, 1951). (It has been shown that the *S* who values an area such as aesthetics highly reports infrequent "aesthetic" words at lower thresholds than an individual who does not value aesthetics highly. However, as Solomon and Howes point out, this may be due to "idiosyncratic" differences in frequency of exposure.) The affective qualities of words influence visual duration threshold *indirectly*, since the frequency of general usage of any word is determined to a considerable extent by the affective tone of that word, with this frequency then acting as a determiner of visual duration threshold (Postman, 1953).

A third position might also be taken concerning the relations between word frequency, word value, and visual duration threshold. Postman (1953) touches upon this last possible interpretation. He says, "Finally, we cannot entirely discount the possibility that familiarity resulting from frequency may be in itself a source of positive value" (p. 68).

It appears, historically, as if the first point of view—that emphasizing word frequency, playing down the influence of value except as it produces idiosyncratic differences in word frequency—has won out. None of the three positions described above have been proven to be untenable, however. Indeed, we have all of the original questions left, plus a few new ones as well. The questions that we see as being testable and requiring further investigation are these:

1. Is word value related, in a general sense, to word frequency? All of the three positions described above would admit the existence of a relation between value and

frequency—they would diverge in opinion with regard to the generality of the relationship. It is true that a knowledge of word frequency allows us to predict visual duration threshold. But why is the frequent word frequent? Perhaps we should go one step beyond frequency to determine whether other attributes of words vary systematically with frequency.

2. If a general relation between value and frequency exists, does it exist because more positively valued words are used more frequently or because one increases positive affect merely by increasing frequency? Or does the dependent–independent relationship depend on the experimental procedure used?

3. If value and frequency are related in a general, not merely an idiosyncratic sense, then one would have to separate out the influences of each on visual duration threshold. Are there significant differences in visual duration threshold between "good" and "bad" words of equal population-wide frequency? Are there significant differences between frequent and infrequent words equal in affective tone?

These are the three problems that we have attempted to deal with in the series of experiments reported below.

Our first concern was to determine whether a general relation exists between word value and word frequency. As we noted above, there is reason to suspect that emotional factors do operate in the building of word frequency. In a large sample of words for which we have actual measures of goodness, the words in the semantic atlas prepared by Jenkins, Russell, and Suci (1958), one finds better than a two to one preponderance of good over bad words. While the words in this atlas are not a random selection of words in the English language, the sources from which the words were selected seem unlikely to be systematically biased in frequency of pleasant as opposed to unpleasant words. One is led to believe that this ratio is characteristic of the entire English language when examining proportions of pleasant to unpleasant words in a sample of 150 words selected at random from the Thorndike-Lorge tables for use in the experiments in Series I, discussed below. Of these 150 words, 95 were on the good end (4.00 or less) of the semantic differential, a ratio of approximately 2:1. It seems highly probable that the sheer number of pleasant words in the English language is far greater than the number of unpleasant words.

It seems equally likely that the frequency with which any specific word is used also depends to a considerable extent on its goodness or badness: hence on its reinforcing qualities. This question is amenable to determination. As Solomon and Howes (1951) state, "In the absence of further data it seems best merely to define word frequency for English words as the frequency of words in the Thorndike-Lorge tables" (pp. 264–265). If affective factors operate in the building of word frequency, then pleasantly toned words should have higher frequencies in the Thorndike-Lorge word count than unpleasantly toned words. Experiment I was designed to determine whether this was the case.

EXPERIMENT I

In the first of this series of experiments we attempted to determine whether the pleasantness of words—in this case measured by the good–bad scale of the semantic differential—was related to word frequency as measured by the Thorndike-Lorge tables.

We selected a sample of words from the Thorndike-Lorge tables. This sample of words consisted of two words selected at random from within every alphabetic category (except X) in the tables. These words were rated on the good–bad dimension of the semantic differential by 24 *Ss*, all freshmen taking an introductory psychology course. The rank order correlation between the goodness and the L count frequency of these words was +.63. Another ran-

dom sample of 50 words was taken from the Thorndike-Lorge word list in the manner described above. These words were rated by 28 *S*s, a different group than those who made the first set of ratings. The rank order correlation between goodness and L count frequency in this list was +.40, again significant to the .01 level of confidence. A third random sample of 50 words was drawn from the Thorndike-Lorge tables. These words were rated by 24 *S*s, none of whom had taken part in previous experiments. The correlation between L count frequency and goodness for this set of words was +.38. All of these three correlations are significant to the .01 level. The results of these experiments, regardless of differences in the magnitude of the correlations, indicate that there is a positive and significant correlation between word frequency and goodness. Value attributes of words are related to word frequency—and not merely in an idiosyncratic way, but in a far more general sense as well.

EXPERIMENT II

Solomon and Howes (1951) list 30 frequent and 30 infrequent words that they used in a study of the relation of word frequency to visual duration threshold. Five frequent and 5 infrequent words had to do with each of the six value areas of the Allport-Vernon Study of Values. Frequent and infrequent words were sometimes, but not always, synonymous. We paired these words, matching the first frequent with the first infrequent word, and so on, through the two lists. Position of presentation within pairs was determined by coin flip. These 30 pairs of words were presented to 34 freshman introductory psychology students, none of whom had taken part in any previous psychological experiments. *S*s were told to "encircle the most pleasantly toned word of each pair." In 26 of the word pairs, the more frequent word was chosen by the majority of the *S*s as the most pleasantly toned. (The 4 exceptions, in which the least frequent word was chosen as the more pleasantly toned, were the word pairs limousine-automobile, economics-assets, orchestra-ensemble, and celestial-heavenly.) A sign test (Guilford, 1956, p. 248) indicates that this preference for the more frequent word in each pair could occur by chance less than one time in a hundred. Again, the more frequent word is rated as the better word.

In the experiment discussed below we attempted to determine whether frequency and goodness were related when nonsense syllables were used as stimulus materials.

EXPERIMENT III

Twenty-two *S*s rated two separate visually represented lists of nonsense syllables for goodness, as measured by the semantic differential. An interval of one week separated the ratings of the two lists. Each list consisted of eight 100%, eight 47–53%, and eight 0% association value (Glaze, 1928) nonsense syllables in random order. Syllables were selected at random from within each block of ten Glaze syllables as listed in the *Handbook of Experimental Psychology* (Stevens, 1951, pp. 540–546) at each of the association values used. Then the eight syllables used in each list at each association level were drawn from this pool of randomly selected syllables. We believed that if frequency of exposure was related to goodness, then the higher the association value of the syllable, the better the syllable should be rated. This seemed likely, since the association value of nonsense syllables appears to be determined largely by the frequency of occurrence of the letter combinations in meaningful words (Underwood

1959). Table 1, showing two sets of syllables and their mean ratings of goodness, indicates this to be the case.

A Mann-Whitney (1947) test of the significance of differences between syllables of 100% and 47–53% association values in List 1 indicates that the 100% list is rated as somewhat better ($P < .10$). The difference in goodness between 47–53% and 0% syllables is significant ($P < .01$); Mann-Whitney tests of List 2 show the 100% and the 47–53% list to differ in goodness ($P < .01$); and the 47–53% list and the 0% list to differ significantly ($P < .05$) as well.

Just as the goodness and the frequency of meaningful words in the English language are related to one another, so also is goodness related to the association value of nonsense syllables; the higher the association value, the better the syllable.

Our data indicate that value and frequency are related. We have moved at least somewhat beyond the empirical fact that word frequency is related to visual duration threshold. We know something else about the stimulus qualities of frequent as opposed to infrequent words; frequent words are not only more frequent but are better as well. Our second question then becomes relevant. Does a manipulation of frequency cause systematic variation in value or do differences in value produce differences in word frequency, or does the independent–dependent status of each of these variables vary depending on the experimental procedures used? Experiment IV bears upon this question.

TABLE 1

RATINGS OF NONSENSE SYLLABLES OF DIFFERING ASSOCIATION VALUES ON THE GOOD–BAD DIMENSION OF THE SEMANTIC DIFFERENTIAL

100% Syllables	Rating	47–53% Syllables	Rating	0% Syllables	Rating
List 1					
BUL	4.72	WOB	4.59	ZEQ	4.91
FES	3.41	KAW	4.33	WUH	4.48
HON	2.11	DUX	4.33	XIW	5.36
JIN	2.98	BOZ	4.00	WUQ	5.26
RAC	3.98	VOK	3.46	QIH	4.87
ROV	3.74	GEY	2.93	GUV	4.39
SUR	2.72	TIZ	3.52	VEC	3.91
WIL	2.46	VOZ	4.30	XUR	5.28
List 2					
DUL	4.27	BEK	3.46	YUF	4.46
BAL	2.35	PIJ	5.17	QIF	4.44
LOV	1.44	WEP	4.71	ZUK	5.04
SAR	3.23	RIQ	4.46	ZIF	4.06
LIK	2.50	YAF	3.96	XAD	4.85
HAV	2.35	LUB	4.15	XUC	5.35
WAT	3.46	GOW	3.75	QUJ	5.00
ROV	3.54	RUV	3.83	GUQ	5.98

EXPERIMENT IV

Fourteen *Ss* rated the goodness of twenty nonsense words. These words were from the list of words used by Solomon and Postman (1952) in their study of the effect of "built in" frequency on visual duration threshold. The words used were: JANDARA, AFWORBU, BIWOJNI, NANSOMA, ENANWAL, IKTITAK, SARICIK, ZABULON, CIVADRA, LOKANTA, KADIRGA, ADAFNAW, BORULCE, NIJARON, ENSHIMI, INDULAM, TAVHANE, UDIBNON, DILIKLI, and MECBURI. Mean goodness scores were obtained. One week later each *S* was tested individually. When the *S* entered the experimental room he was given the following instructions (from Solomon & Postman, 1952):

> This is an experiment concerning the effectiveness of repetition in learning to pronounce strange words correctly. It has a direct bearing on the problem of reading words in a foreign language, as compared to hearing the words spoken. In addition, we are interested in knowing whether the relative effectiveness of the two kinds of learning methods depends on general reading ability.
>
> We are going to give you a deck of cards On each card is printed a strange word. We would like you to look at each card carefully and then pronounce the word in the way it would be pronounced if it were a word in the

English language. Proceed steadily from card to card, turning over each one after you have finished with it. Go right through the deck and then stop unless you have serious doubts about your pronunciations.

The experimenter (E) handed the S a pack of 90 3″ ×5″ cards. In this pack were the words listed above. The first five words above were presented 10 times apiece, the next five, 5 times apiece, the next five, 2 times apiece, and the last five, once each. The cards were shuffled thoroughly after each S completed this task, so that they were in a relatively random order.

All responses were taped in order to lend realism to the expressed purpose of this new situation. Since Ss had participated in another experiment on the pronounceability of words, it seems likely that (although the Ss had seen the words before while rating them on the semantic differential) they were misled as to the purpose of this experiment. Each S rated the nonsense words a second time following this pronounceability test.

Figure 1 indicates the mean preexposure and postexposture ratings for words of each frequency of exposure. While the differences between ratings prior to exposure were not significant, a Kruskal and Wallis H test (1952) shows that words of differing frequencies of exposure do differ significantly in the postexposure test ($P < .01$, $H = 7.83$).

In this experimental situation a manipulation of word frequency alters Ss' evaluations of word goodness. One can, however, think of instances where the reverse might well occur. The most obvious example of this is in verbal conditioning, where the differential reinforcement of a class of words produces systematic variation in the frequency with which this class of words is emitted. We would suspect that in this case frequency of emission (and hence of reinforcement) alters the affective quality of the words. Reinforcement might then produce further variation of frequency.

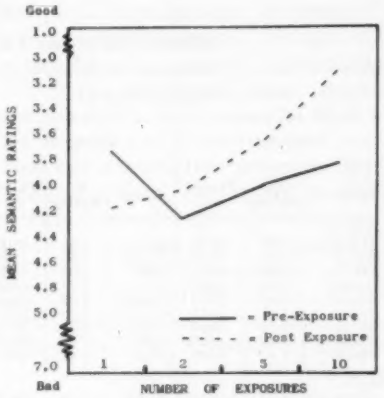

FIG. 1. Ratings of nonsense words before and after word frequency was manipulated.

No matter whether frequency of exposure alters value or vice versa, or both, one thing is evident—there is a relation between word frequency and Ss' ratings of word goodness. Since this is so, our third question also becomes relevant. Since value and frequency are generally related, it became necessary to experiment with samples of words matched in frequency, varying in goodness, plus samples of words varying in frequency, matched in goodness, in order to separate out the influence of value and of frequency on visual duration threshold. This was attempted in Experiment V.

EXPERIMENT V

We had various groups of Ss rate the goodness of a large number of words on the semantic differential. These groups rated 60 words of a Thorndike-Lorge G count of 12, 13, or 14, three random samples of 50 words each (described in Experiment I), and 39 words, one sample of 30 words plus

TABLE 2

WORDS USED IN THE TACHISTOSCOPIC EXPERIMENT

Words of Matched Goodness						Words of Matched Frequency					
Frequent			Infrequent			Good			Bad		
Word	*SD* Rating	Threshold	Word	*SD* Rating	Threshold	Word	G Count	Threshold	Word	G Count	Threshold
Sample 1											
Think	2.34	.0373	Income	2.52	.0500	Rally	12	.1092	Stammer	12	.1538
Butter	1.97	.0692	Caress	1.97	.1085	Glisten	13	.0777	Corrupt	12	.0985
Doctor	1.60	.0369	Ivory	1.68	.0262	Prosper	13	.1046	Defect	14	.0815
Health	1.97	.0492	Piano	1.97	.0546	Ivory	13	.0262	Hinder	13	.0969
Sleep	2.22	.0362	Radiant	2.08	.0662	Ginger	13	.0446	Wretch	12	.1285
Sweet	1.93	.0631	Patriot	1.90	.0992	Alert	14	.0392	Sneer	12	.0785
Mother	1.73	.0315	Vision	1.57	.0546	Deputy	13	.0446	Scandal	13	.1123
House	2.46	.0323	Agile	2.74	.0623	Vehicle	13	.0492	Dismal	13	.1162
Flower	1.87	.0369	Bible	1.87	.0462	Caress	12	.1085	Penalty	13	.0469
Sample 2											
Mountain	2.73	.0687	Bodice	2.74	.0965	Bodice	2	.0965	Slime	4	.0713
Water	2.43	.0217	Dough	2.47	.0757	Candy	32	.0409	Beggar	29	.1087
Window	2.68	.0400	Gradual	2.63	.0965	Engine	A	.0565	Bitter	A	.1704
Green	2.93	.0296	Elegant	3.00	.0852	Brave	A	.0739	Anger	A	.0417
Smooth	2.57	.0522	Zenith	2.62	.0330	Statue	32	.0748	Divorce	29	.0948
Father	2.36	.0800	Jelly	2.49	.0800	Cushion	21	.0900	Inferior	19	.1948
Pretty	2.13	.0661	Refresh	2.09	.0661	Garment	40	.1043	Grief	45	.0687
Street	2.77	.0496	Lenient	2.77	.1522	Church	AA	.0530	Danger	AA	.0583

9 words which happened to appear on the other lists, from Jenkins' semantic atlas. From these words, and from Jenkins' atlas, we selected 17 pairs of words matched in frequency, varying in goodness, and 17 pairs of words matched in goodness, varying in frequency. (Some words among those matched in goodness were rated by Jenkins' *S*s and by our own. Although the rank order correlation of goodness ratings between our own and Jenkins' sample of *S*s was +.98, for a group of 30 of these words, there was some variation in ratings between the two samples. Whenever this occurred, we used the mean goodness rating assigned by our *S*s.) The first 9 words from each list were presented to the first group of 26 *S*s; the last 8 to a second group of 23 *S*s. The four lists of words, divided as they were presented to two sample populations, are presented in Table 2.

The words were put on slides in the form of bold faced capital letters. The words were presented to *S*s by means of a Revere 888 tachistoscope. Each *S* sat 18 feet from the screen on which the words were projected. The testing room was fully illuminated while the tachistoscope was used at the smallest possible lens opening. The ratio of room illumination to screen illumination was 1:1.08. The words were presented in random order with the order remaining constant in all presentations to any specific *S* but with the order varying between *S*s. A modified method of limits was used in which each of the 36 words was presented once at each of the following speeds: 1/100, 1/50, 1/25, 1/5, 1/2, and 1 second. Under the conditions of presentation used in this experiment, only one word was recognized at 1/100 of a second while all words were recognized at 1 second.

Each *S*'s mean visual duration threshold for frequent as opposed to infrequent and good as opposed to bad words was obtained. The *S*s in the first sample ($N = 26$), exposed to the first group of words, differed significantly in mean thresholds for good vs.

bad words ($t = 3.56$, $P < .01$, using the formula for correlated data), with good words reported at lower thresholds. These *S*s also differed in mean thresholds for frequent vs. infrequent words ($t = 2.33$, $P < .05$), with frequent words reported at lower thresholds. The *S*s in the second sample ($N = 23$), exposed to the second group of words, differed significantly in mean thresholds for good vs. bad words ($t = 3.25$, $P < .01$), with good words reported at lower thresholds, and in threshold for frequent vs. infrequent words ($t = 3.82$, $P < .01$), with frequent words reported at lower thresholds.

DISCUSSION

We attempted to obtain data relevant to three specific questions. Let us look at these questions, along with the relevant data.

1. Is value related, in a general sense, to word frequency? Our data is unequivocally affirmative. Whether dealing with samples of words in the English language, with frequent vs. infrequent words in various Allport-Vernon value areas, or with nonsense syllables, frequency, as measured by the Thorndike-Lorge tables and by association value, and positive value, as measured by the good-bad scale of the semantic differential, are significantly related in a general sense.

2. If this general relation exists, does it exist because more positively valued words are used more frequently or because one increases positive value merely by increasing frequency? We found that a manipulation of frequency produces systematic variation in *S*s' ratings of the goodness of words. It seems highly probable, however, that the reverse may also be true—that by increasing the goodness of words, one also increases frequency. At this point it seems likely that the independent-dependent relation between goodness and frequency would depend on the experimental situation. We would not hazard a guess as to the effect of each on the other in general speech, although a relation between frequency and goodness has been shown to exist.

The data gathered in the first three experiments suggest that other attributes of the stimulus vary systematically as one varies frequency. Perhaps this is true for other forms of learning also. The fourth experiment would suggest the possibility that a manipulation of the frequency of occurrence of a stimulus and a contiguous response, in itself and without any regard to drive reduction, changes the reinforcing qualities of the stimulus. Perhaps we should look beyond the influence of frequency on response probability and ask *why* frequency changes response probability. Is it the difference in association value (and hence, in frequency of occurrence) that causes 100% association value syllables to be learned more rapidly than 0% syllables? Or is it the variation in affective quality accompanying differing frequencies of occurrence? Since the relation between goodness and frequency is not perfect, one can obtain syllables that vary from 0 to 100% in association value, yet are equal in semantic rating—this should provide a test for our question.

These results seem to have considerable social consequence. If word frequency and word goodness are related, and if a manipulation of the frequency of occurrence of a word produces systematic variation in word value, then one wonders whether ideas and personal values are manipulable in the same fashion. If so, then perhaps those elements of society that believe in censorship would find support in these experiments for one of their major assumptions, viz., that frequent exposure to an idea or a set of behaviors—violence, drinking alcoholic beverages, or what have you—lessens the "evil" associated with the idea of behavior. The validity of their second major assumption—that if we make a thing evil enough, people won't do this thing—is, of course, a very different problem.

If the more frequent becomes demonstrably better merely because it is

more frequent, one can well understand the values of advertising. If it is generally true that the most frequent words in the English language are also the most good, the logical conclusion that this would lead to is that ads would be more effective if written at an even lower vocabulary level than is currently used. "Oh, brave new world. . . ."

3. If value and frequency are related, what are the influences of each on visual duration threshold when the other variable is held constant? Words equal in frequency, varying in value, are correctly reported by *S*s at different mean thresholds, with good words being correctly identified at significantly lower thresholds than are bad words. Words equal in goodness, varying in frequency, are reported by *S*s at different mean thresholds, with frequent words being correctly identified at significantly lower thresholds.

While it has been generally accepted that values have some secondary effect on visual duration thresholds, it is also generally accepted that the influence of values on threshold is small, producing idiosyncratic variations in frequency and hence influencing thresholds in this way. Our results indicate that values influence threshold even when frequency of exposure is held constant, and that this influence is approximately equal to that of frequency with value held constant, using these lists of words.

Since word goodness, independent from frequency, does significantly influence visual duration threshold, we are left with the problem of "Why?" The terms "perceptual selectivity" and "perceptual defense" are descriptive, not explanatory. We have no answers here, although we would like to discuss several possible explanations.

One interpretation of the fact that good words are perceived at lower threshold than bad words of the same frequency might be called the "simple" guess theory. Good words appear to be about twice as common as bad words in the English language. Specific good words are likely to be more frequently used than specific bad words. It seems, then, that good words occur between three and four times as frequently as bad words in terms of usage. Even though the frequency of occurrence of the stimuli are equaled, by matching good and bad words for Thorndike-Lorge frequency, response probabilities probably are still not equal. If an *S* sees a word in a tachistoscope but identifies only a "g—— d——" he is probably three or four times more likely, on the basis of responses made in the past, to fill in the gaps with a good word, such as "gradual," rather than a bad word, such as "goudge." This simple guess explanation may be partially correct but is not too satisfying to us, since an examination of our *S*s' responses to words on the good–bad lists indicates that there seems to be a rather strong likelihood that when *S*s guess at but miss a bad word, they respond with another bad word—incorrect but still bad, while when they respond incorrectly to the tachistoscope presentation of a good word, the incorrect response is usually another good word. This finding runs contrary to the idea that perceptual defense is a significant variable but also suggests that something more systematic than mere guessing is involved in errors and in correct responses to tachistoscopically presented words.

A second explanation for the differences in *S*s' mean visual duration thresholds might be called the "complex" guess theory. This explanation is derived from Bruner and Postman's (1949) hypothesis theory of perception and cognition. According to Postman (1951) hypotheses are, "in the most general sense, expectancies or predispositions of the organism which serve to select, organize, and transform the stimulus information that comes from the environment" (p. 249). The

strength or availability of certain hypotheses is stronger than that of others. This strength is a function of the following factors: frequency of past confirmation of the hypothesis, number of alternative hypotheses, motivational relevance, and cognitive support. The first two factors—frequency of past confirmation and number of alternative hypotheses—fit well into the simple guess theory, while the last factor, that of cognitive support, may not be relevant to this discussion. The factor of motivational relevance might possibly, however, produce differential predictions between the two explanatory devices that we have labeled simple and complex guessing. As Postman (1951) puts it: "The stronger the motivational support for a given hypothesis, the less of appropriate information is required to confirm it" (p. 255). We would limit this statement somewhat by tying it only to the effect of reward or reinforcement on the number of cues required for the confirmation of a hypothesis. It may be that the differential reinforcement of various stimuli causes those stimuli (as well as those responses) that have most frequently led to reinforcement in the past to be more readily observable or available in the present. The availability of a specific response may well depend on the frequency that this response has been rewarded in the past—but perhaps because that aspect of the stimulus situation that has called forth the rewarded response in the past has become more salient—has become more of a "figure" as opposed to the other stimuli becoming more like the "ground." To paraphrase our quotation from Postman given above: The more reinforcing or rewarding a word is, the less cues are needed for the recognition of this word. We are advancing the proposition that reinforcement not only influences the probability that a given response will be emitted, but also that reinforcement causes that aspect of the stimulus situation that is salient at the time that the rewarded response is made to become more likely to be salient at the next presentation of the stimulus. If this is true, then even if word frequencies of good and bad words are equal and even if habitual modes of response (such as guessing good words) are controlled, the stimulus qualities of rewarding or reinforcing words would be such that less information is needed before the word is correctly recognized than would be the case for words which were presented with equal frequency but were not equally reinforcing. Something that might be called perceptual sensitivity might then be said to exist. This sensitivity would not, however, be a unique perceptual process, but would instead depend on those old standbys in all forms of learning: frequency and reinforcement, a little intermingled to be sure, but still observably present.

We have no way of testing out the "why" question with these data, although the types of wrong guesses made in our study incline us away from the simple guess theory. We believe that the data presented in this study are valuable chiefly in reviving the concept of perceptual selectivity, not in solving the problem of why what is called perceptual selectivity exists. By systematically varying frequency of exposure and frequency of reinforcement and nonreinforcement of nonsense syllables in future experiments, we hope to eventually be able to deal with the question of why good words are perceived at lower mean thresholds than are bad words.

SUMMARY

Three positions may be taken concerning the relation between word value and word frequency, and the relation of each of these to visual duration thresh-

old. Two of these points of view would suggest a general relation to exist between word frequency and word value. Three experiments demonstrate the existence of this relationship. The more frequent a word or a nonsense syllable occurs in the English language, the better it is likely to be rated on the good–bad scale of the semantic differential. Since this relation does exist, we attempted to test out, in Experiment IV, one possible explanation of *why* the relation exists. A manipulation of the frequency of nonsense words produces systematic variation in the rated goodness of these nonsense words. Words that are frequent are also very likely to be rated as good. Hence, to separate out the influence of word value and of word frequency on visual duration threshold, we tachistoscopically presented lists of words matched in frequency, varying in goodness; matched in goodness, varying in frequency. Subjects reported the good words at significantly lower thresholds than the matched bad words; the frequent words at significantly lower thresholds than matched infrequent words. Both frequency and value appear to operate in producing differential visual duration thresholds for words. Possible explanations and certain implications of these findings were discussed.

REFERENCES

BRUNER, J. S., & POSTMAN, L. Perception, cognition, and behavior. *J. Pers.*, 1949, **18**, 12–31.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* (3rd ed.) New York: McGraw-Hill, 1956.

JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles of 360 words. *Amer. J. Psychol.*, 1958, **71**, 688–699.

KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one criterion variance analysis. *J. Amer. Statist. Ass.*, 1952, **47**, 583–621.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

POSTMAN, L. Towards a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 242–272.

POSTMAN, L. The experimental analyses of motivational factors in perception. In J. S. Brown (Ed.), *Current theory and research in motivation.* Lincoln: Univer. Nebraska Press, 1953. Pp. 59–108.

POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271–284.

SOLOMON, R. L., & HOWES, D. H. Word frequency, word value, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

SOLOMON, R. L., & POSTMAN, L. Usage as a determinant of visual duration thresholds of words. *J. exp. Psychol.*, 1952, **43**, 195–201.

STEVENS, S. S. (Ed.) *Handbook of experimental psychology.* New York: Wiley, 1951.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, 1944.

UNDERWOOD, B. J. Verbal learning and the educative process. *Harvard educ. Rev.*, 1959, **29**, 107–117.

(Received September 15, 1959)

CLINICAL INFERENCE AND COGNITIVE THEORY

THEODORE R. SARBIN
RONALD TAFT
DANIEL E. BAILEY

The original theory of cognition developed here adds a new dimension to the consideration of diagnosis construction, and personality assessment and description. August, 1960.

PLANS AND THE STRUCTURE OF BEHAVIOR

GEORGE A. MILLER
EUGENE GALANTER
KARL H. PRIBRAM

This discussion of the organism as an information processing system incorporates the important notion of feedback. The TOTE concept developed here is a fundamentally different explanation of behavior from that provided by the reflex arc.

PSYCHOLOGY OF PERCEPTION

WILLIAM N. DEMBER

Thoroughly up-to-date in research coverage, clear, precise and literate in style, this book is designed to introduce the student to current problems and developments in research on perception.

HOLT, RINEHART and WINSTON, Inc., New York

CONTEMPORARY PSYCHOLOGY

. . . . APA's journal of book reviews

- Critical reviews of the most recent books in psychology and related areas
- The Editor's page: comments on news in the publishing world, on criticism, on literary style
- A letters-to-the-editor section
- Instructional Media: reviews and comments on films, TV programs, workbooks, teaching machines

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth St., N.W.
Washington 6, D. C.